

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-267700
(P2000-267700A)

(43) 公開日 平成12年9月29日 (2000.9.29)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 1 0 L 19/02		G 1 0 L 9/14	P 5 D 0 4 5
19/08			G 5 J 0 6 4
11/06		9/00	C 9 A 0 0 1
H 0 3 M 7/30		H 0 3 M 7/30	B

審査請求 有 請求項の数 7 O L (全 24 頁)

(21) 出願番号 特願平11-72062

(22) 出願日 平成11年3月17日 (1999.3.17)

(71) 出願人 399004577

株式会社ワイ・アール・ビー高機能移動体
通信研究所
神奈川県横須賀市光の丘3番2号

(72) 発明者 佐々木 誠司

神奈川県横須賀市光の丘3番2号 株式会
社ワイ・アール・ビー高機能移動体通信研
究所内

(74) 代理人 100106459

弁理士 高橋 英生 (外3名)

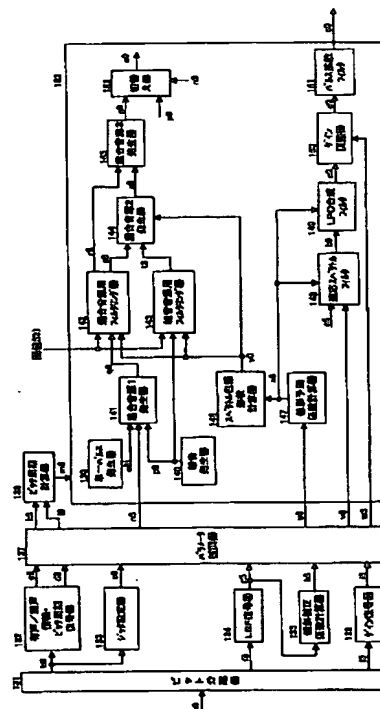
最終頁に続く

(54) 【発明の名称】 音声符号化復号方法および装置

(57) 【要約】

【課題】 音質の劣下を伴わずに更なる低ビットレート化を実現する。

【解決手段】 復号器において、スペクトル包絡情報により算出される周波数軸上のスペクトル包絡値と予め定めた閾値とを比較して、該スペクトル包絡値が閾値以上になる周波数領域を有声領域、その他の領域を無声領域とし、それぞれに適した音源を用いて音源信号を作ることにより、付加的な周波数帯別の音源切替え情報の伝送を必要とせずに音質の向上を図る。符号化器において、非周期的ピッチを有するフレームでは、ピッチ周期の発生度数に偏りがあることに着目し、非周期的ピッチの周期に対して、その度数の大小に応じた不均一量子化を行った結果を、無声状態および周期的ピッチの量子化結果と共にまとめて一つの符号として伝送することにより伝送ビット数を削減し低ビットレート化を図る。



1

【特許請求の範囲】

【請求項 1】 線形予測分析・合成方式の音声符号化器によって音声信号が符号化処理された出力である音声情報ビット列から音声信号を再生する音声復号方法であって、

前記音声情報ビット列に含まれるスペクトル包絡情報、有声／無声識別情報、ピッチ周期情報およびゲイン情報を分離、復号し、

前記有声／無声識別情報が有声を示すときには、前記スペクトル包絡情報により算出される周波数軸上のスペクトル包絡値と予め定めた閾値とを比較して、該スペクトル包絡値が前記閾値以上になる周波数領域を有声領域、その他の領域を無声領域とし、有声領域の音源信号として前記ピッチ周期情報に基づき発生されるピッチパルスを用い、無声領域の音源信号として前記ピッチパルスと白色雑音を所定の割合で混合した信号を用い、前記有声領域の音源信号および前記無声領域の音源信号を加算した結果を音源信号とし、

前記有声／無声識別情報が無声を示すときには、白色雑音を音源信号とし、

該音源信号に対し前記スペクトル包絡情報および前記ゲイン情報を付加して再生音声を作成することを特徴とする音声復号方法。

【請求項 2】 線形予測分析・合成方式の音声符号化器によって音声信号が符号化処理された出力である音声情報ビット列から音声信号を再生する音声復号方法であって、

前記音声情報ビット列に含まれるスペクトル包絡情報、有声／無声識別情報、ピッチ周期情報およびゲイン情報を分離、復号し、

前記有声／無声識別情報が有声を示す場合は、有声強度情報に 1.0 をセットし、有声／無声識別情報が無声を示す場合は、有声強度情報に 0 をセットし、

前記スペクトル包絡情報、前記ピッチ周期情報、前記ゲイン情報および該有声強度情報をピッチ周期に同期して線形補間し、

該補間された有声強度情報に応じた割合で、前記補間されたピッチ周期情報に基づき発生されるピッチパルスと白色雑音を混合して第 1 の混合音源信号とし、

前記補間されたスペクトル包絡情報により算出される周波数軸上のスペクトル包絡値と予め定めた閾値とを比較して、該スペクトル包絡値が閾値以上になる周波数領域を有声領域、その他の領域を無声領域とし、有声領域の音源信号として前記第 1 の混合音源信号を用い、無声領域の音源信号として前記第 1 の混合音源信号と白色雑音とを所定の割合で混合した信号を用い、前記有声領域の音源信号および前記無声領域の音源信号を加算した結果を第 2 の混合音源信号とし、

該第 2 の混合音源信号に対し前記補間されたスペクトル包絡情報および前記補間されたゲイン情報を付加して再

2

生音声を作成することを特徴とする音声復号方法。

【請求項 3】 標本化され、予め定められた時間長の音声符号化フレームに分割された入力音声信号から、有声／無声識別情報、ピッチ周期情報、周期的ピッチか非周期的ピッチかを示す非周期ピッチ情報を抽出して、符号化する音声符号化方法であって、

前記非周期ピッチ情報が周期的ピッチを示す音声符号化フレームでは、前記ピッチ周期情報を第 1 の所定のレベル数で量子化して、これを周期的ピッチ情報とし、

10 前記非周期ピッチ情報が非周期的ピッチを示す音声符号化フレームでは、それぞれのピッチ範囲に対しその発生度数の大小に応じた量子化レベルの割り当てを行い、第 2 の所定のレベル数で量子化して、これを非周期的ピッチ情報とし、

前記有声／無声識別情報が無声を示す状態に 1 つの符号語を割り当て、前記有声／無声識別情報が有声を示す状態として、前記周期的ピッチ情報に前記第 1 の所定のレベル数に対応する個数の符号語を割り当て、前記非周期的ピッチ情報に前記第 2 の所定のレベル数に対応する個数の符号語を割り当て、これらをまとめて所定のビット数を有する符号語として符号化することを特徴とする音声符号化方法。

【請求項 4】 前記符号語のビット数は 7 ビットであり、前記有声／無声識別情報が無声を示す状態に全てのビットが 0 (又は 1) の符号語を割り当て、前記非周期的ピッチ情報に 7 ビット中、1 または 2 ビットが 1 (又は 0) である符号語を割り当て、その他の符号語に前記周期的ピッチ情報を割り当てることを特徴とする前記請求項 3 記載の音声符号化方法。

30 【請求項 5】 前記請求項 3 に記載された音声符号化方法と、前記請求項 1 あるいは 2 に記載された音声復号方法とからなる音声符号化復号方法。

【請求項 6】 所定のサンプル周波数で標本化され、量子化された音声サンプルを入力し、予め定められた時間長の音声符号化フレーム毎に所定数の音声サンプルを出力するフレーム化器と、

該 1 フレーム分の音声サンプルのレベル情報である RMS 値の対数を計算し、その結果である対数 RMS 値を出力するゲイン計算器と、

40 該対数 RMS 値を線形量子化し、その結果である量子化後の対数 RMS 値を出力する第 1 の量子化器と、

前記 1 フレーム分の音声サンプルに対し線形予測分析を行い、スペクトル包絡情報である所定次数の線形予測係数を出力する線形予測分析器と、

該線形予測係数を L S F (Line Spectrum Frequencies) 係数に変換して出力する L S F 係数計算器と、

該 L S F 係数を量子化し、その結果である L S F パラメータインデックスを出力する第 2 の量子化器と、

50 前記 1 フレーム分の音声サンプルを所定のカットオフ周波数でフィルタリングし帯域制限された入力信号を出力

3

するローパスフィルタと、
 該帯域制限された入力信号から正規化自己相関関数計算に基づきピッチ周期を抽出し、ピッチ周期および正規化自己相関関数の最大値を出力するピッチ検出器と、
 該ピッチ周期を対数変換した後、第1の所定のレベル数で線形量子化し、その結果であるピッチ周期インデックスを出力する第3の量子化器と、
 前記正規化自己相関関数の最大値を入力し、所定の閾値より小さければ非周期フラグをONにセット、そうでなければOFFにセットして、非周期フラグを出力する非周期フラグ発生器と、
 前記線形予測係数を係数として用いて前記1フレーム分の音声サンプルからスペクトル包絡情報を除去し、その結果である残差信号を出力するLPC分析フィルタと、
 該残差信号を入力し、ピーキネス値を計算し、ピーキネス値を出力するピーキネス計算器と、
 該ピーキネス計算器の値により、前記正規化自己相関関数の最大値の値を補正して補正された正規化自己相関関数の最大値を出力する相関関数補正器と、
 該補正された正規化自己相関関数の最大値が所定の閾値以下であれば無声、そうでなければ有声と判定し、その結果である有声／無声フラグを出力する有声／無声判定器と、
 前記非周期フラグが非周期を示しているフレームの前記ピッチ周期について、第2の所定のレベル数で不均一量子化し、非周期的ピッチインデックスを出力する非周期ピッチインデックス生成器と、
 前記有声／無声フラグ、前記非周期フラグ、前記ピッチ周期インデックス、および前記非周期的ピッチインデックスを入力し、これらを所定のビット数で符号化した周期／非周期ピッチ・有声／無声情報コードを出力する周期／非周期ピッチおよび有声／無声情報コード生成器と、
 前記量子化後の対数RMS値、前記LSFパラメータインデックス、および前記周期／非周期ピッチ・有声／無声情報コードを入力し、1フレーム毎にビットパッキングを行い音声情報ビット列を出力するビットパッキング器とを備えた音声符号化装置。
 【請求項7】 前記請求項6記載の音声符号化装置により生成された1フレーム毎の音声情報ビット列を各パラメータ毎に分離し、周期／非周期ピッチ・有声／無声情報コード、量子化後の対数RMS値およびLSFパラメータインデックスを出力するビット分離器と、
 前記周期／非周期ピッチ・有声／無声情報コードを入力し、現フレームの状態が無声の場合は、ピッチ周期を所定の値にセットし、有声強度を0にセットして出力し、周期的および非周期的の場合は、ピッチ周期を符号化の規則に基づき復号処理して出力し、有声強度を1.0にセットして出力する有声／無声情報・ピッチ周期復号器と、

4

前記周期／非周期ピッチ・有声／無声情報コードを入力し、現フレームが無声または非周期的を示す場合は、ジッタ値を所定の値にセットして出力し、周期的を示す場合は、ジッタ値を0にセットして出力するジッタ設定器と、
 前記LSFパラメータインデックスから前記所定の次数のLSF係数を復号して出力するLSF復号器と、
 該LSF係数から傾斜補正係数を計算し出力する傾斜補正係数計算器と、
 10 前記量子化後の対数RMS値を復号し、ゲインを出力するゲイン復号器と、
 前記ピッチ周期、前記有声強度、前記ジッタ値、前記LSF係数、前記傾斜補正係数および前記ゲインを、それぞれピッチ周期に同期して線形補間し、補間後のピッチ周期、補間後の有声強度、補間後のジッタ値、補間後のLSF係数、補間後の傾斜補正係数および補間後のゲインを出力するパラメータ補間器と、
 該補間後のピッチ周期および補間後のジッタ値を入力し、補間後のピッチ周期にジッタを付加した後、整数値
 20 に変換されたピッチ周期（以後、整数ピッチ周期という）を出力するピッチ周期計算器と、
 該整数ピッチ周期に同期して該整数ピッチ周期分の再生音声を復号し出力する1ピッチ波形復号器とを備え、
 該1ピッチ波形復号器は、
 前記整数ピッチ周期期間内に単一パルス信号を出力する単一パルス発生器と、
 前記整数ピッチ周期の長さを持つ白色雑音を出力する雑音発生器と、
 前記補間後の有声強度に基づき、前記単一パルス信号と該白色雑音とを合成して第1の混合音源信号を出力する第1の混合音源発生器と、
 前記補間後のLSF係数から線形予測係数を計算する線形予測係数計算器と、
 30 該線形予測係数から再生音声のスペクトル包絡形状情報を求め出力するスペクトル包絡形状計算器と、
 前記スペクトル包絡形状情報の値と所定の閾値とを比較し、スペクトル包絡形状情報の値が該閾値以上の周波数領域を有声領域、それ以外の周波数領域を無声領域として、前記第1の混合音源情報のDFT係数のうち前記無
 40 声領域のものを0にした第1のDFT係数列、および、
 前記混合音源情報のDFT係数のうち前記有声領域のものを0とした第2のDFT係数列を出力する混合音源用フィルタリング器と、
 前記白色雑音のDFT係数のうち前記有声領域のものと0としたDFT係数列を出力する雑音音源用フィルタリング器と、
 前記混合音源用フィルタリング器からの前記第2のDFT係数列と前記雑音音源用フィルタリング器からのDFT係数列とを所定の割合で混合して、その結果である
 50 DFT係数列を出力する第2の混合音源発生器と、

該第2の混合音源発生器からのDFT係数列と前記混合音源用フィルタからの前記第1のDFT係数列とを加算した後、逆離散フーリエ変換を行い、混合音源信号として出力する第3の混合音源発生器と、

前記補間後の有声強度を入力し、それが0のときは前記白色雑音を選択し、その他のときは前記第3の混合音源発生器の出力である混合音源信号を選択して混合音源信号として出力する切替え器と、

前記線形予測係数に帯域幅拡張処理を施したものを係数とする適応極／零フィルタと、前記補間後の傾斜補正係数を係数とするスペクトル傾斜補正フィルタの従属接続であり、前記混合音源信号をフィルタリングしてスペクトルが改善された音源信号を出力する適応スペクトルエンハンスメントフィルタと、

前記線形予測係数を係数として用いる全極型フィルタであり、該スペクトルが改善された音源信号に対してスペクトル包絡情報を付加して、スペクトル包絡情報が付加された信号を出力するLPC合成フィルタと、

該スペクトル包絡情報が付加された信号に対し、前記ゲインを用いてゲイン調整を行い、再生音声信号を出力するゲイン調整器と、

該再生音声信号に対し、パルス拡散処理を施し、パルス拡散処理された再生音声信号を出力するパルス拡散フィルタとを備える音声復号装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、音声信号を低ビットレートで符号化処理及び復号処理する音声符号化方法とその装置に関するものである。

【0002】

【従来の技術】低ビットレート音声符号化方式として、2.4kbps LPC (Linear Predictive Coding: 線形予測符号化) 方式と2.4kbps MELP (Mixed Excitation Linear Prediction: 混合音源線形予測) 方式が知られている。これらは共に米国連邦政府標準の音声符号化方式であり、前者はFS-1015 (FSはFederal Standard) として標準化されており、後者は、FS-1015の音質改良版として1996年に新たに選定、標準化された。

【0003】この2.4kbps LPC方式および2.4kbps MELP方式に関して、次のような参考文献がある。

[1] FEDERAL STANDARD 1015, "ANALOG TO DIGITAL CONVERSION OF VOICE BY 2400 BIT/SECOND LINEAR PREDICTIVE CODING", November 28, 1984

[2] Federal Information Processing Standards Publication, "Analog to digital conversion of voice by 2,400 Bit/Second Mixed Excitation Linear Prediction", May 28, 1998 Draft

[3] L. Supplee, R. Cohn, J. Collura and A. McCree, "MELP: The new federal standard at 2400 bps", Proc. ICASSP, pp. 1591-1594, 1997

[4] A. McCree and T. Barnwell III, "A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding", IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, VOL. 3, NO. 4, pp. 242-250, July 1995

[5] D. Thomson and D. Prezas, "SELECTIVE MODELING OF THE LPC RESIDUAL DURING UNVOICED FRAMES: WHITE NOISE OR PULSE EXCITATION", Proc. ICASSP, pp. 3087-3090, 1986

[6] 佐々木、三宅, "線形予測分析・合成方式の復号器", 特許第2711737号公報

【0004】最初に、2.4kbps LPC方式の原理について図9および図10を用いて説明する(処理の詳細については、参考文献[1]を参照されたい)。図9はLPC方式の音声符号化器の構成を示すブロック図である。フレーム化器(11)は、100-3600Hzで帯域制限された後、8kHzで標準化され、少なくとも12ビットの精度で量子化された入力音声サンプル(a1)を蓄えるバッファであり、1音声符号化フレーム(22.5ms)毎に音声サンプル(180サンプル)を取り込み、音声符号化処理部へ(b1)として出力する。以下では、1音声符号化フレーム毎に実行される処理について説明する。

【0005】プリエンファシス器(12)は、(b1)を高域強調処理し、高域強調処理された信号(c1)を出力する。線形予測分析器(13)は、(c1)をDurbin-Levinson法を用いて線形予測分析し、スペクトル包絡情報である10次の反射係数(d1)を出力する。量子化器1(14)は(d1)を各次数毎にスカラー量子化し、その結果である計41ビットを(e1)を誤り訂正符号化/ビットパッキング器(19)へ出力する。各次数の反射係数に対するビット配分は表1に示す。RMS (Root Mean Square: 実効値) 計算器(15)は高域強調処理された信号(c1)のレベル情報であるRMS値を計算し、RMS値(f1)を出力する。量子化器2(16)は(f1)を5ビットで量子化し、その結果である(g1)を誤り訂正符号化/ビットパッキング器(19)へ出力する。

【0006】ピッチ検出/音響分類器(17)は、前記フレーム化器11の出力(b1)を入力し、ピッチ周期(20~156サンプル(51~400Hzに対応)の範囲をとる)および音響分類情報(有声/無声/過渡部の識別情報)を抽出し、それぞれ(h1)、(i1)として出力する。量子化器3(18)は(h1)および(i1)をまとめて7ビットで量子化し、その結果(j1)を誤り訂正符号化/ビットパッキング器(19)へ出力する。ここでの量子化方法(7ビットの符号(128種類の符号語)へのピッチ情報、音響分類情報の割り当て方)は、7ビットが全て0の符号語および7ビット中1ビットのみが1となる符号語を無声に割り当て、7ビットが全て1の符号語および7ビット中1ビットのみが0となる符号語を過渡部に割り当てる。その他の符号語は有声用としてピッチ周期情報に割り当てられる。誤り訂正符号化/ビットパッキング器(19)は量子化されたそれぞれの情報(e1)、(g1)、(j1)を54ビット/フレーム

にパッキングし、音声符号化情報フレームを構成し、1フレーム毎に54ビットを(k1)として出力する。音声情報ビット列(k1)は、無線通信の場合、変調器、無線機を通り、受信側に伝送される。

【0007】表1に1フレーム当たりのビット配分を示す。同表から分かるように、誤り訂正符号化/ビットパッキング器(19)では、そのフレームの音響的分類が有声でないならば(つまり無声または過渡部であるなら

表1 2.4kbpsLPC方式のビット配分

パラメータ	有声フレーム	無声フレーム
反射係数(1次)	5	5
反射係数(2次)	5	5
反射係数(3次)	5	5
反射係数(4次)	5	5
反射係数(5次)	4	—
反射係数(6次)	4	—
反射係数(7次)	4	—
反射係数(8次)	4	—
反射係数(9次)	3	—
反射係数(10次)	2	—
ピッチおよび音響分類情報	7	7
RMS	5	—
誤り保護	—	20
同期ビット	1	1
未使用	—	1
合計ビット/22.5msフレーム	54	64

【0009】次に図10を用いてLPC音声復号器の構成について説明する。ビット分離/誤り訂正復号器(21)は1フレーム毎に受信した54ビットの音声情報ビット列(a2)を各パラメータ毎に分離すると共に、そのフレームが無声または過渡部の場合には前記該当ビットに対して誤り訂正復号処理を施す。そして、その結果であるピッチ/音響分類情報ビット(b2)、10次の反射係数情報ビット(e2)およびRMS情報ビット(g2)を出力する。ピッチ/音響分類情報復号器(22)は前記ピッチ/音響分類情報ビット(b2)を復号し、ピッチ周期(c2)および音響分類情報(d2)を出力する。反射係数復号器(23)は前記10次の反射係数情報ビット(e2)を復号し、10次の反射係数(f2)を出力する。RMS復号器(24)は前記RMS情報ビット(g2)を復号し、RMS情報(h2)を出力する。パラメータ補間器(25)は、再生音声の品質を向上するため、各パラメータ(c2)、(d2)、(f2)、(h2)をそれぞれ補間処理し、その結果である(i2)、(j2)、(o2)、(r2)を出力する。

【0010】次に音源信号(m2)は以下のようにして作られる。音響分類切替器(28)は、前記補間された音響分類情報(j2)が有声を示す時は、パルス音源発生器(26)が発生するピッチ周期(i2)に同期したパルス音源(k2)を選択し、音響分類情報(j2)が無声音声部を示す時には雑音発生器(27)が発生する白色雑音(12)を選択するように動作する。また、音響分類情報(j2)が過渡部を示す時には、そのフレーム内の有声部分に対してはパルス音源(k2)、無声部分に対しては白色雑音(擬似ランダム音源)(12)を選択するように動作する。ここでフレーム内にお

ば、5～10次の反射係数を送る代わりに誤り訂正符号(20ビット)を送る。無声または過渡部の場合に誤り保護される情報は、RMS情報の上位4ビット、1～4次の反射係数情報である。また、各フレーム毎に1ビットの同期ビットが付加される。

【0008】

【表1】

ける有声部分と無声部分の境界はパラメータ補間器(25)で決定される。また、ここで使用されるパルス音源(k2)を作るためのピッチ周期情報(i2)は、隣接した有声音フレームのものを使用する。音響分類切替器(28)の出力が音源信号(m2)となる。

【0011】LPC合成フィルタ(30)は、線形予測係数(p2)を係数として用いる全極型フィルタであり、音源信号(m2)に対しスペクトル包絡情報を付加して、その結果である信号(n2)を出力する。ここで、スペクトル包絡情報である線形予測係数(p2)は、線形予測係数計算器(29)により前記反射係数(o2)から計算される。また、LPC合成フィルタ(30)は、有声に対しては10次の線形予測係数(p2)を用いる10次の全極型フィルタとして構成され、無声に対しては4次の線形予測係数(p2)を用いる4次の全極型フィルタとして構成される。ゲイン調整器(31)は前記LPC合成フィルタ(30)の出力(n2)に対し前記RMS情報(r2)を用いてゲイン調整を行い、(q2)を出力する。最後にデエンファシス器(32)は、(q2)に対し、前述のプリエンファシス器(12)と逆の処理を行い再生音声(s2)を出力する。

【0012】このようなLPC方式の問題点を以下に示す(参考文献[4])。

問題点A：LPC方式では、全周波数帯域に渡り、フレーム毎に有声/無声/過渡部を切り替えている。しかし、自然音声の音源信号は、小さな周波数帯域に分けて観測すると有声の性質を持つ帯域と、無声の性質を持つ帯域がある。従って、LPC方式において有声と決定さ

れたフレームでは、雑音で駆動すべき成分をパルスで駆動してしまうため、buzz音（ブンブンとうなるような音）になる。これは、高いほうの周波数で顕著になる。問題点B：無声から有声に変化する過渡部では、非周期性パルスを有する音源信号となる場合があるが、LPC方式の過渡部フレームでは、非周期的パルス音源を表現できない。そのため、トーンの雑音が生じる。このように、LPC方式では、buzz音、トーンの雑音の発生により再生音声は聞きづらい音質（機械的な音質）となってしまうという問題点がある。

【0013】次に、上記のようなLPC方式の問題点を解決し、音質改良を図った方式であるMELP方式について説明する（参考文献[2]-[4]）。まず、MELP方式ではどのような方法で音質改善しているかについて図11を用いて説明する。同図（a）に示すように、自然音声の周波数軸上で帯域に分けて見ると、白で示されている周期的パルス成分が支配的な帯域（有声部）と、黒で示されている雑音成分が支配的な帯域（無声部）とが存在する。上述のようにLPCボコーダで再生音が機械

的になる主な原因は、同図（b）に示すように、周波数帯域全体にわたり、有声フレームでは周期的パルス成分で、無声音フレームでは雑音成分で音源を表現しているからである（過渡部フレームにおいては、フレームを時間的に有声と無声に分けている。）。この問題を解決するため、MELP方式では、同図（c）に示すように、1フレーム内で5つの周波数帯域（サブバンド）毎に有声／無声を切替えることにより混合音源を適用している。この手法は、上記LPC方式の問題点Aを解決し、再生音声におけるbuzz音を低減する効果がある。また、上記LPC方式の問題点Bを解決するため、非周期的パルス情報を抽出、伝送し、復号器側で非周期的パルス音源を生成する機能を有している。その他、再生音声の音質改善のため、適応スペクトルエンハンスメントフィルタ、パルス拡散フィルタおよびハーモニクス振幅情報の利用という手法を取り入れている。表2にMELP方式で使用される各手法の効果をまとめる。

【0014】

【表2】

表2 MELP方式で使用される各手法の効果

手法	効果
① 混合音源	周波数帯域毎に有声／無声判定を行っているので、buzz音を低減できる。
② 非周期的パルス	過渡部、破裂音で生じる不規則な（非周期的な）声門パルスを表現することにより、トーンの雑音を低減できる。
③ 適応スペクトルエンハンスメントフィルタ	ホルマントの共振を鋭くし、自然音声のホルマントに対する近似度を改善することにより再生音声の自然性を向上させる。
④ パルス拡散フィルタ	自然音声の声門パルス波形に対するパルス音源波形の近似度を改善することにより再生音声の自然性を向上させる。
⑤ ハーモニクス振幅	スペクトル表現をより正確にすることにより、鼻音の品質、話者識別の性能および広帯域雑音がある時の母音の品質を向上できる。

【0015】次に、2.4kbps MELP方式の構成について図12および図13を用いて説明する（処理の詳細については、参考文献[2]を参照されたい）。図12はMELP音声符号化器の構成を示すブロック図である。フレーム化器(41)は、100-3800Hzで帯域制限された後、8kHzで標準化され、少なくとも12ビットの精度で量子化された入力音声サンプル(a3)を蓄えるバッファであり、1音声符号化フレーム（22.5ms）毎に音声サンプル（180サンプル）を取り込み、音声符号化処理部へ(b3)として出力する。以下では1音声符号化フレーム毎に実行される処理について説明する。

【0016】ゲイン計算器(42)は(b3)のレベル情報であるRMS値の対数を計算し、その結果である(c3)を出力する。この処理はフレームの前半と後半について行われる。すなわち、1フレーム当たり2つのRMS値の対数を(c3)として出力する。量子化器1(43)は、(c3)を前半のものについて3ビット、後半のものについて5ビットで線形量子化し、その結果である(d3)を誤り訂正符号化／ビットパッキング器(70)へ出力する。線形予測分析器(44)は、(b3)をDurbin-Levinson法を用いて線形予測分

析し、スペクトル包絡情報である10次の線形予測係数(e3)を出力する。LSF係数計算器(45)は、10次の線形予測係数(e3)を10次のLSF（Line Spectrum Frequencies）係数(f3)に変換する。LSF係数は、線形予測係数と等価な特徴パラメータであるが、それに比べ、量子化特性、補間特性に優れるため、最近の殆ど音声符号化方式に採用されている。量子化器2(46)は10次のLSF係数(f3)を段数4の多段ベクトル量子化により25ビットで量子化し、(g3)を誤り訂正符号化／ビットパッキング器(70)へ出力する。

【0017】ピッチ検出器(54)は、前記フレーム化器41の出力(b3)の1kHz以下の信号成分に対して整数ピッチ周期を求めた後、この整数ピッチ周期と、(b3)をLPF（ローパスフィルタ）(55)により500Hz以下に帯域制限した信号(q3)とを用いて小数ピッチ周期を求め、(r3)として出力する。ピッチ周期は正規化自己相関関数が最大となる遅延量として与えられるが、この時の正規化自己相関関数の最大値(o3)も出力される。正規化自己相関関数の最大値の大きさは、入力信号(b3)の周期性の強さを表す情報であり、非周期フラグ発生器(56)（後で説明す

る)で用いられる。また正規化自己相関関数の最大値(o3)は、相関関数補正器(53)(後で説明する)で補正された後、誤り訂正符号化/ビットパッキング器(70)における全帯域の有声/無声判定に用いられる。そこでは、補正後の正規化自己相関関数の最大値(n3)が閾値(=0.6)以下であれば無声、そうでなければ有声と判定される。量子化器3(57)は前記ピッチ検出器54からの小数ピッチ周期(r3)を入力し対数変換した後、99レベルで線形量子化し、その結果である(s3)を誤り訂正符号化/ビットパッキング器(70)へ出力する。

【0018】4つのBPF(バンドパスフィルタ)(58)、(59)、(60)および(61)は、前記フレーム化器41の出力(b3)をそれぞれ500~1000Hz、1000~2000Hz、2000~3000Hz、3000~4000Hzで帯域制限し、(t3)、(u3)、(v3)および(w3)を出力する。4つの自己相関計算器(62)、(63)、(64)および(65)は、それぞれ(t3)、(u3)、(v3)および(w3)に対し、小数ピッチ周期(r3)に対応する遅延量における正規化自己相関関数を計算し、(x3)、(y3)、(z3)および(a4)として出力する。次に4つの有声/無声フラグ発生器(66)、(67)、(68)および(69)は、それぞれ(x3)、(y3)、(z3)および(a4)に対し、閾値(=0.6)以下であれば無声、そうでなければ有声と判定し、有声/無声

$$\text{ピーキネス値 } p = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N e_n^2}}{\frac{1}{N} \sum_{n=1}^N |e_n|}$$

ここで、Nは1フレーム中のサンプル数、 e_n は残差信号である。上式(1)の分子は分母に比べ大きな値の影響を受けやすいので、pは残差信号中に大きなスパイクが存在する時に大きな値となる。従って、ピーキネス値が大きいほど、そのフレームが破裂音フレーム又は非周期的なパルス列を有するフレームである可能性が大きくなる(破裂音のフレームでは、部分的にスパイク(鋭いピーク)を持つが、その他の部分は、白色雑音に近い性質の信号になっているため)。

【0020】相関関数補正器(53)は、ピーキネス値(m3)の値により、正規化自己相関関数の最大値(o3)および有声/無声フラグ(b4)、(c4)、(d4)および(e4)の値を補正する。ピーキネス値(m3)が1.34より大きければ、正規化自己相関関数の最大値(o3)を1.0(有声を示す)にセットする。また、ピーキネス値(m3)が1.6より大きければ正規化自己相関関数の最大値(o3)を1.0(有声を示す)にセットすると共に有声/無声フラグ(b4)、(c4)、(d4)および(e4)を有声にセットする。補正後の正規化自己相関関数の最大値は(n3)として出力され、補正後の有声/無声フラグ(b4)、(c4)、(d4)および(e4)は帯域毎の有声情報(f4)として出力される。

【0021】前述のように、破裂音のフレームでは、部

を示すフラグ(1ビット)を(b4)、(c4)、(d4)および(e4)として相関関数補正器(53)へ出力する。これらの各帯域の有声/無声フラグ(b4)、(c4)、(d4)および(e4)は、復号器において混合音源を生成するのに用いられる。非周期フラグ発生器(56)は、正規化自己相関関数の最大値(o3)を入力し、閾値(=0.5)より小さければ非周期フラグをONにセット、そうでなければOFFにセットして、非周期フラグ(1ビット)(p3)を誤り訂正符号化/ビットパッキング器(70)へ出力する。非周期フラグ(p3)は復号器において過渡部、破裂音の音源を表現するための非周期性パルスを生成するのに用いられる。

【0019】LPC分析フィルタ1(51)は10次の線形予測係数(e3)を係数として用いる全零型フィルタであり、入力音声(b3)からスペクトル包絡情報を除去し、その結果である残差信号(13)を出力する。ピーキネス計算器(56)は、残差信号(13)を入力し、ピーキネス値を計算し(m3)として出力する。ピーキネス値とは、信号中にピークをもつパルスの成分(スパイク)が存在する可能性を表すパラメータであり、上記参考文献[5]より次式で与えられる。

【数1】

(1)

分的にスパイク(鋭いピーク)を持つが、その他の部分は、白色雑音に近い性質の信号になっているため、正規化自己相関関数が0.5より小さな値(このとき非周期フラグがONにセットされる)となる可能性が大きい。そこで、ピーキネス値により破裂音フレームを検出して正規化自己相関関数を1.0に補正すれば、その後の誤り訂正符号化/ビットパッキング器(70)における全帯域の有声/無声判定において有声と判断され、復号の際に非周期パルスを音源として用いることにより、破裂音フレームの音質は改善される。また、過渡部に多くみられる非周期的なパルス列を持つフレームも、同じ理由により音質は改善される。

【0022】次にハーモニクス情報の検出について説明する。線形予測係数計算器(47)は、量子化器2(46)の出力である量子化後のLSF係数(g3)を線形予測係数に変換し、量子化後の線形予測係数(h3)を出力する。LPC分析フィルタ2(48)は、(h3)を係数として入力信号(b3)からスペクトル包絡成分を除去し、残差信号(i3)を出力する。ハーモニクス検出器(49)は(i3)における10次のハーモニクス(基本ピッチ周波数の高調波成分)の振幅を抽出し、その結果である(j3)を出力する。量子化器4(50)は、(j3)を8ビットでベクトル量子化し、その

インデックス(k3)を誤り訂正符号化／ビットパッキング器(70)へ出力する。ハーモニクス振幅情報は、残差信号(i3)に残っているスペクトル包絡情報に相当する。従って、ハーモニクス振幅情報を送ることにより、復号時に入力信号のスペクトル表現をより正確に表現することができ、鼻音の品質、話者識別の性能および広帯域雑音がある時の母音の品質を向上させることができる(表2)。

【0023】誤り訂正符号化／ビットパッキング器(70)は、前述したように補正後の正規化自己相関関数の最大値(n3)が閾値(=0.6)以下であれば無声フレーム、そうでなければ有声フレームと設定し、表3に示すビット配分で音声情報ビット列を構成し、1フレーム毎に54ビットを(g4)として出力する。音声情報ビット列(g4)は、無線通信の場合、変調器、無線機を通り、受信側に伝送さ

表3 2.4kbps MELP方式のビット配分

パラメータ	有声フレーム	無声フレーム
LSFパラメータ	25	25
ハーモニクス振幅	8	—
ゲイン2回／フレーム	8	8
ピッチ、全体の有声／無声情報	7	7
帯域毎の有声性	4	—
非周期フラグ	1	—
誤り保護	—	13
同期ビット	1	1
合計ビット／22.5ms フレーム	54	54

【0025】次に図13を用いてMELP音声復号器の構成について説明する。ビット分離／誤り訂正復号器(81)は1フレーム毎に受信した54ビットの音声情報ビット列(a5)の中からピッチ、全体の有声／無声情報を取り出し、それが無声フレームを示す場合には誤り保護該当ビットに対して誤り訂正復号処理を施す。また、ピッチ、全体の有声／無声情報がイレースを示す場合には、各パラメータを前フレームのもので置換処理する。そして、分離された各パラメータの情報ビットとして、ピッチ、全体の有声／無声情報(b5)、非周期フラグ(d5)、ハーモニクス振幅インデックス(e5)、帯域毎の有声性(g5)、LSFパラメータインデックス(j5)、およびゲイン情報(m5)を出力する。ここで、帯域毎の有声性(g5)は、各サブバンド(0~500Hz、500~1000Hz、1000~2000Hz、2000~3000Hz、3000~4000Hz)の有声性を示す5ビットのフラグであり、0~500Hzの有声性については、ピッチ、全体の有声／無声情報から取り出された全体の有声／無声情報を用いる。

【0026】ピッチ復号器(82)は、ピッチ、全体の有声／無声情報が有声を示す場合にはピッチ周期を復号し、無声を示す場合はピッチ周期として50.0をセットして復号されたピッチ周期(c5)を出力する。ジッタ設定器(102)は、非周期フラグ(d5)を入力し、非周期フラグがONを示すならばジッタ値を0.25、OFFを示すならばジッタ値を0にセットし、(g6)を出力する。ここで、上記の有声

れる。表3において、ピッチ、全体の有声／無声情報が7ビットで量子化されているが、その方法は次の通りである。7ビットの符号(128種類)うち、7ビットが全て0の符号語および7ビット中1ビットのみが1となる符号語を無声に割り当て、7ビット中2ビットが1となる符号語をイレージャ(消失)に割り当てる。その他の符号語は有声用としてピッチ周期情報(量子化器3(57)の出力(s3))に割り当てられる。また、同表から分かるように、そのフレームが無声ならばハーモニクス振幅(k3)、帯域毎の有声性(f4)および非周期フラグ(p3)を送る代わりに、聴感上重要なビットに対し誤り訂正を施し、その誤り訂正符号(13ビット)を送る。また、各フレーム毎に1ビットの同期ビットが付加される。

【0024】

【表3】

／無声情報が無声を示す場合は、ジッタ値(g6)は0.25にセットされる。ハーモニクス復号器(83)は、ハーモニクス振幅のインデックス(e5)から10次のハーモニクス振幅(f5)を復号し出力する。パルス音源用フィルタ係数計算器(84)は、帯域毎の有声性(g5)を入力し、有声を示しているサブバンドのゲインを1.0、無声を示しているサブバンドのゲインを0にするようなFIRフィルタの係数(h5)を計算し、出力する。また、雑音音源用フィルタ係数計算器(85)は帯域毎の有声性(g5)を入力し、有声を示しているサブバンドのゲインを0、無声を示しているサブバンドのゲインを1.0にするようなFIRフィルタの係数(i5)を計算し、出力する。LSF復号器(87)は、LSFパラメータインデックス(j5)から10次のLSF係数(k5)を復号し、出力する。傾斜補正係数計算器(86)は、10次のLSF係数(k5)から傾斜補正係数(l5)を計算する。ゲイン復号器(88)は、ゲイン情報(m5)を復号し、ゲイン(n5)を出力する。

【0027】パラメータ補間器(89)は、各パラメータ(c5)、(g6)、(f5)、(h5)、(i5)、(l5)、(k5)および(n5)についてそれぞれピッチ周期に同期して線形補間し、(o5)、(p5)、(r5)、(s5)、(t5)、(u5)、(v5)および(w5)を出力する。ここでの線形補間処理は、次式により実施される。

補間後のパラメータ=現フレームのパラメータ×int + 前フレームのパラメータ×(1.0-int)

ここで、現フレームのパラメータは(c5)、(g6)、(f5)、(h5)、(i5)、(l5)、(k5)および(n5)のそれぞれに対応し、補間後のパラメータは(o5)、(p5)、(r5)、(s5)、(t5)、(u5)、(v5)および(w5)のそれぞれに対応する。前フレームのパラメータは、前フレームにおける(c5)、(g6)、(f5)、(h5)、(i5)、(l5)、(k5)および(n5)を保持しておくことにより与えられる。intは補間係数であり、次式で求める。

$$\text{int} = \text{to} / 180$$

ここで、180は音声復号フレーム長(22.5ms)当たりのサンプル数、toは、復号フレームにおける1ピッチ周期の開始点であり、1ピッチ周期分の再生音声は復号される毎にそのピッチ周期が加算されることにより更新される。toが180を超えるとそのフレームの復号処理が終了したことになり、toから180が減算される。

【0028】ピッチ周期計算器(90)は、補間されたピッチ周期(o5)およびジッタ値(p5)を入力し、ピッチ周期(q5)を次式により計算する。

$$\text{ピッチ周期}(q5) = \text{ピッチ周期}(o5) \times (1.0 - \text{ジッタ値}(p5) \times \text{乱数値})$$

ここで、乱数値は-1.0~1.0の範囲の値をとる。上式より無声または非周期的フレームではジッタ値が0.25にセットされているのでジッタが付加され、周期的フレームではジッタ値が0にセットされているのでジッタは付加されない。但し、ジッタ値はピッチ毎に補間処理されているので、0~0.25の範囲をとるため中間的なピッチ区間も存在する。このように非周期フラグに基づき非周期ピッチ(ジッタが付加されたピッチ)を発生することは、表2の に示したように過渡部、破裂音で生じる不規則な(非周期的な)声門パルス表現することにより、トーンの雑音を低減する効果がある。

【0029】ピッチ周期(q5)は整数値に変換された後、1ピッチ波形復号器(101)に入力される。1ピッチ波形復号器(101)は、ピッチ周期(q5)毎の再生音声(f6)を復号し出力する。従って、このブロックに含まれる全てのブロックはピッチ周期(q5)を入力し、それに同期して動作する。パルス音源発生器(91)は、補間されたハーモニクス振幅(r5)を入力し、そのハーモニクス情報が付加された単一パルスを有するパルス音源(x5)を発生する。このパルス音源(x5)はピッチ周期(q5)に1パルス発生される。パルスフィルタ(92)は、補間されたパルスフィルタ用係数(s5)を係数とするFIRフィルタであり、パルス音源(x5)に対し有声のサブバンドのみを有効にするようにフィルタリングし、(y5)を出力する。雑音発生器(94)は、白色雑音(a6)を発生する。雑音フィルタ(93)は、補間された雑音フィルタ用係数(t5)を係数とするFIRフィルタであり、雑音音源(a6)に対し無声のサブバンドのみを有効にするようにフィルタリングし、(z5)を出力する。混合音源発生器(95)は(y5)および(z5)を加算し、混合音源(b6)を発生する。この混合音源は、表2の

に示したように周波数帯毎に有声/無声音源を切り替えることによりbuzz音を低減する効果がある。

【0030】線形予測係数計算器(98)は補間された10次のL S F係数(v5)から線形予測係数(h6)を計算する。適応スペクトルエンハンスメントフィルタ(96)は、線形予測係数(h6)に帯域幅拡張処理を施したものを係数とする適応極/零フィルタであり、表2の に示した通り、ホルマントの共振を鋭くし、自然音声のホルマントに対する近似度を改善することにより再生音声の自然性を向上させる。さらに、補間された傾斜補正係数(u5)を用いてスペクトルの傾きを補正して音のこもりを低減し、その結果である音源信号(c6)を出力する。L P C合成フィルタ(97)は、線形予測係数(h6)を係数として用いる全極型フィルタであり、音源信号(c6)に対しスペクトル包絡情報を付加して、その結果である信号(d6)を出力する。ゲイン調整器(99)は(d6)に対しゲイン情報(w5)を用いてゲイン調整を行い、(e6)を出力する。パルス拡散フィルタ(100)は、自然音声の声門パルス波形に対するパルス音源波形の近似度を改善するためのフィルタであり、(e6)をフィルタリングして自然性が改善された再生音声(f6)を出力する。このパルス拡散フィルタの効果は表2の に示す通りである。以上により、MELP方式では、L P C方式に比べ、同ビットレート(2.4 kbps)において自然性、了解性の高い再生音声を提供することができる。

【0031】さらに、MELP方式のように帯域毎の有声性情報を伝送する必要なしに、上述のL P C方式の問題点Aを解決しようとした線形予測分析・合成方式の復号器が提案されている(参考文献[6])。この提案されている線形予測分析・合成方式の復号器の内容を以下に示す。線形予測分析・合成方式の符号器によって音声信号が分析符号化された入力ディジタル信号に含まれる線形予測係数、有声無声識別信号、音源強度情報およびピッチ周期情報の各パラメータを分離して出力する分離回路と、前記ピッチ周期情報によって制御されたピッチパルスを出力するピッチパルス発生器と、白色雑音を出力する雑音発生器と、前記ピッチパルス発生器の出力と前記雑音発生器の出力と駆動音源として前記線形予測係数に従って復号化した音声信号を出力する合成フィルタとを備えた線形予測分析・合成方式の復号器において、前記分離回路からの線形予測係数と有声無声識別信号および音源強度情報とが入力され、有声音のホルマントの合成による周波数軸上のスペクトル包絡線を求めて予め定めた閾値と比較し、該スペクトル包絡線のレベルが閾値より大きい周波数領域を表すピッチ成分関数信号と、該スペクトル包絡線のレベルが閾値より小さい周波数領域を表す雑音成分関数信号とを出力する処理制御回路と、前記ピッチパルス発生器の出力と前記ピッチ成分関数信号とを乗算して前記閾値より大きい周波数領域のピッチパルスを出力する第1の出力制御回路と、前記雑音発生

器からの白色雑音と前記雑音成分関数信号とを乗算して前記閾値より小さい周波数領域の白色雑音を出力する第2の出力制御回路と、前記第1の出力制御回路と第2の出力制御回路とを合成して前記合成フィルタに対する駆動音源として出力する合成器とを備えたことを特徴とする線形予測分析・合成方式の復号器。しかし、この手段では、前述のLPC方式の問題点であるbuzz音は軽減できるものの、再生音声の音質が雑音的になるという欠点がある（この原因については後に述べる）。

【0032】

【発明が解決しようとする課題】移動体通信の爆発的普及により、ユーザ収容数の増大が必要となっており、周波数資源の更なる有効利用が課題となっている。音声符号化方式の更なる低ビットレート化は、この課題を解決するための必須の技術課題の1つである。そこで、本発明は、2.4 kbpsより低いビットレートで、上述のLPC方式の問題点AおよびBを解決することのできる音声符号化復号方法および装置を提供することを目的としている。また、MELP方式のように帯域毎の有声性情報や非周期フラグを伝送する必要なしに、それらと同様の効果が得られる音声符号化復号方法および装置を提供することを目的としている。

【0033】

【課題を解決するための手段】上記目的を達成するために、本発明の音声復号方法は、線形予測分析・合成方式の音声符号化器によって音声信号が符号化処理された出力である音声情報ビット列から音声信号を再生する音声復号方法であって、前記音声情報ビット列に含まれるスペクトル包絡情報、有声／無声識別情報、ピッチ周期情報およびゲイン情報を分離、復号し、前記有声／無声識別情報が有声を示すときには、前記スペクトル包絡情報により算出される周波数軸上のスペクトル包絡値と予め定めた閾値とを比較して、該スペクトル包絡値が前記閾値以上になる周波数領域を有声領域、その他の領域を無声領域とし、有声領域の音源信号として前記ピッチ周期情報に基づき発生されるピッチパルスを用い、無声領域の音源信号として前記ピッチパルスと白色雑音を所定の割合で混合した信号を用い、前記有声領域の音源信号および前記無声領域の音源信号を加算した結果を音源信号とし、前記有声／無声識別情報が無声を示すときには、白色雑音を音源信号とし、該音源信号に対し前記スペクトル包絡情報および前記ゲイン情報を付加して再生音声の生成することとを特徴とするものである。これにより、付加的な情報ビットを伝送することなく、前述したLPC方式の問題点Aを解決することができる。

【0034】また、本発明の他の音声復号方法は、線形予測分析・合成方式の音声符号化器によって音声信号が符号化処理された出力である音声情報ビット列から音声信号を再生する音声復号方法であって、前記音声情報ビット列に含まれるスペクトル包絡情報、有声／無声識別

情報、ピッチ周期情報およびゲイン情報を分離、復号し、前記有声／無声識別情報が有声を示す場合は、有声強度情報に1.0をセットし、有声／無声識別情報が無声を示す場合は、有声強度情報に0をセットし、前記スペクトル包絡情報、前記ピッチ周期情報、前記ゲイン情報および該有声強度情報をピッチ周期に同期して線形補間し、該補間された有声強度情報に応じた割合で、前記補間されたピッチ周期情報に基づき発生されるピッチパルスと白色雑音を混合して第1の混合音源信号とし、前記補間されたスペクトル包絡情報により算出される周波数軸上のスペクトル包絡値と予め定めた閾値とを比較して、該スペクトル包絡値が閾値以上になる周波数領域を有声領域、その他の領域を無声領域とし、有声領域の音源信号として前記第1の混合音源信号を用い、無声領域の音源信号として前記第1の混合音源信号と白色雑音とを所定の割合で混合した信号を用い、前記有声領域の音源信号および前記無声領域の音源信号を加算した結果を第2の混合音源信号とし、該第2の混合音源信号に対し前記補間されたスペクトル包絡情報および前記補間されたゲイン情報を付加して再生音声の生成することとを特徴とするものである。これにより、付加的な情報ビットを伝送することなく、上述したLPC方式の問題点Aを解決することができる。

【0035】さらに、本発明の音声符号化方法は、標準化され、予め定められた時間長の音声符号化フレームに分割された入力音声信号から、有声／無声識別情報、ピッチ周期情報、周期的ピッチか非周期的ピッチかを示す非周期ピッチ情報を抽出して、符号化する音声符号化方法であって、前記非周期ピッチ情報が周期的ピッチを示す音声符号化フレームでは、前記ピッチ周期情報を第1の所定のレベル数で量子化して、これを周期的ピッチ情報とし、前記非周期ピッチ情報が非周期的ピッチを示す音声符号化フレームでは、それぞれのピッチ範囲に対しその発生度数の大小に応じた量子化レベルの割り当てを行い、第2の所定のレベル数で量子化して、これを非周期的ピッチ情報とし、前記有声／無声識別情報が無声を示す状態に1つの符号語を割り当て、前記有声／無声識別情報が有声を示す状態として、前記周期的ピッチ情報に前記第1の所定のレベル数に対応する個数の符号語を割り当て、前記非周期的ピッチ情報に前記第2の所定のレベル数に対応する個数の符号語を割り当て、これらをまとめて所定のビット数を有する符号語として符号化することとを特徴とするものである。さらにまた、前記符号語のビット数は7ビットであり、前記有声／無声識別情報が無声を示す状態に全てのビットが0（又は1）の符号語を割り当て、前記非周期的ピッチ情報に7ビット中、1または2ビットが1（又は0）である符号語を割り当て、その他の符号語に前記周期的ピッチ情報を割り当てるものである。これにより、付加的な情報ビットを伝送することなく、上述したLPC方式の問題点Bを解

決することができる。また、低ビットレートの音声符号化を実現することができる。

【0036】さらにまた、本発明の音声符号化復号方法は、上述した音声復号方法のいずれかと上述した音声符号化方法とを組み合わせたものである。これにより、付加的な情報ビットを伝送することなく、上述したLPC方式における問題点AおよびBを解決することができる。

【0037】さらにまた、本発明の音声符号化装置は、所定のサンプル周波数で標準化され、量子化された音声サンプルを入力し、予め定められた時間長の音声符号化フレーム毎に所定数の音声サンプルを出力するフレーム化器と、該1フレーム分の音声サンプルのレベル情報であるRMS値の対数を計算し、その結果である対数RMS値を出力するゲイン計算器と、該対数RMS値を線形量子化し、その結果である量子化後の対数RMS値を出力する第1の量子化器と、前記1フレーム分の音声サンプルに対し線形予測分析を行い、スペクトル包絡情報である所定次数の線形予測係数を出力する線形予測分析器と、該線形予測係数をLSF (Line Spectrum Frequencies) 係数に変換して出力するLSF係数計算器と、該LSF係数を量子化し、その結果であるLSFパラメータインデックスを出力する第2の量子化器と、前記1フレーム分の音声サンプルを所定のカットオフ周波数でフィルタリングし帯域制限された入力信号を出力するローパスフィルタと、該帯域制限された入力信号から正規化自己相関関数計算に基づきピッチ周期を抽出し、ピッチ周期および正規化自己相関関数の最大値を出力するピッチ検出器と、該ピッチ周期を対数変換した後、第1の所定のレベル数で線形量子化し、その結果であるピッチ周期インデックスを出力する第3の量子化器と、前記正規化自己相関関数の最大値を入力し、所定の閾値より小さければ非周期フラグをONにセット、そうでなければOFFにセットして、非周期フラグを出力する非周期フラグ発生器と、前記線形予測係数を係数として用いて前記1フレーム分の音声サンプルからスペクトル包絡情報を除去し、その結果である残差信号を出力するLPC分析フィルタと、該残差信号を入力し、ピーキネス値を計算し、ピーキネス値を出力するピーキネス計算器と、該ピーキネス計算器の値により、前記正規化自己相関関数の最大値の値を補正して補正された正規化自己相関関数の最大値を出力する相関関数補正器と、該補正された正規化自己相関関数の最大値が所定の閾値以下であれば無声、そうでなければ有声と判定し、その結果である有声/無声フラグを出力する有声/無声判定器と、前記非周期フラグが非周期を示しているフレームの前記ピッチ周期について、第2の所定のレベル数で不均一量子化し、非周期的ピッチインデックスを出力する非周期ピッチインデックス生成器と、前記有声/無声フラグ、前記非周期フラグ、前記ピッチ周期インデックス、および前記非周期的

ピッチインデックスを入力し、これらを所定のビット数で符号化した周期/非周期ピッチ・有声/無声情報コードを出力する周期/非周期ピッチおよび有声/無声情報コード生成器と、前記量子化後の対数RMS値、前記LSFパラメータインデックス、および前記周期/非周期ピッチ・有声/無声情報コードを入力し、1フレーム毎にビットパッキングを行い音声情報ビット列を出力するビットパッキング器とを備えたものである。

【0038】さらにまた、本発明の音声復号装置は、前記音声符号化装置により生成された1フレーム毎の音声情報ビット列を各パラメータ毎に分離し、周期/非周期ピッチ・有声/無声情報コード、量子化後の対数RMS値およびLSFパラメータインデックスを出力するビット分離器と、前記周期/非周期ピッチ・有声/無声情報コードを入力し、現フレームの状態が無声の場合は、ピッチ周期を所定の値にセットし、有声強度を0にセットして出力し、周期的および非周期的の場合は、ピッチ周期を符号化の規則に基づき復号処理して出力し、有声強度を1.0にセットして出力する有声/無声情報・ピッチ周期復号器と、前記周期/非周期ピッチ・有声/無声情報コードを入力し、現フレームが無声または非周期的を示す場合は、ジッタ値を所定の値にセットして出力し、周期的を示す場合は、ジッタ値を0にセットして出力するジッタ設定器と、前記LSFパラメータインデックスから前記所定の次数のLSF係数を復号して出力するLSF復号器と、該LSF係数から傾斜補正係数を計算し出力する傾斜補正係数計算器と、前記量子化後の対数RMS値を復号し、ゲインを出力するゲイン復号器と、前記ピッチ周期、前記有声強度、前記ジッタ値、前記LSF係数、前記傾斜補正係数および前記ゲインを、それぞれピッチ周期に同期して線形補間し、補間後のピッチ周期、補間後の有声強度、補間後のジッタ値、補間後のLSF係数、補間後の傾斜補正係数および補間後のゲインを出力するパラメータ補間器と、該補間後のピッチ周期および補間後のジッタ値を入力し、補間後のピッチ周期にジッタを付加した後、整数値に変換されたピッチ周期（以後、整数ピッチ周期という）を出力するピッチ周期計算器と、該整数ピッチ周期に同期して該整数ピッチ周期分の再生音声を復号し出力する1ピッチ波形復号器とを備え、該1ピッチ波形復号器は、前記整数ピッチ周期期間内に単一パルス信号を出力する単一パルス発生器と、前記整数ピッチ周期の長さを持つ白色雑音を出力する雑音発生器と、前記補間後の有声強度に基づき、前記単一パルス信号と該白色雑音とを合成して第1の混合音源信号を出力する第1の混合音源発生器と、前記補間後のLSF係数から線形予測係数を計算する線形予測係数計算器と、該線形予測係数から再生音声のスペクトル包絡形状情報を求め出力するスペクトル包絡形状計算器と、前記スペクトル包絡形状情報の値と所定の閾値とを比較し、スペクトル包絡形状情報の値が該閾値以上の周

波数領域を有声領域、それ以外の周波数領域を無声領域として、前記第1の混合音源情報のDFT係数のうち前記無声領域のものを0にした第1のDFT係数列、および、前記混合音源情報のDFT係数のうち前記有声領域のものを0とした第2のDFT係数列を出力する混合音源用フィルタリング器と、前記白色雑音のDFT係数のうち前記有声領域のものと0としたDFT係数列を出力する雑音音源用フィルタリング器と、前記混合音源用フィルタリング器からの前記第2のDFT係数列と前記雑音音源用フィルタリング器からのDFT係数列とを所定の割合で混合して、その結果であるDFT係数列を出力する第2の混合音源発生器と、該第2の混合音源発生器からのDFT係数列と前記混合音源用フィルタからの前記第1のDFT係数列とを加算した後、逆離散フーリエ変換を行い、混合音源信号として出力する第3の混合音源発生器と、前記補間後の有声強度を入力し、それが0のときは前記白色雑音を選択し、その他のときは前記第3の混合音源発生器の出力である混合音源信号を選択して混合音源信号として出力する切替器と、前記線形予測係数に帯域幅拡張処理を施したものを係数とする適応極／零フィルタと、前記補間後の傾斜補正係数を係数とするスペクトル傾斜補正フィルタの従属接続であり、前記混合音源信号をフィルタリングしてスペクトルが改善された音源信号を出力する適応スペクトルエンハンスメントフィルタと、前記線形予測係数を係数として用いる全極型フィルタであり、該スペクトルが改善された音源信号に対してスペクトル包絡情報を付加して、スペクトル包絡情報が付加された信号を出力するLPC合成フィルタと、該スペクトル包絡情報が付加された信号に対し、前記ゲインを用いてゲイン調整を行い、再生音声信号を出力するゲイン調整器と、該再生音声信号に対し、パルス拡散処理を施し、パルス拡散処理された再生音声信号を出力するパルス拡散フィルタとを備えるものである。

【0039】

【発明の実施の形態】本発明の音声符号化復号方法および装置の一実施の形態について、図1～8を用いて詳しく説明する。なお、以下では、具体的な数値を用いて説明するが、本発明は以下の説明に用いた数値以外の数値を用いても実施することができる点に注意されたい。図1は、本発明の音声符号化復号方法が適用された音声符号化器の一構成例のブロック図である。この図において、フレーム化器(111)は、100-3800Hzで帯域制限された後、8kHzで標準化され、少なくとも12ビットの精度で量子化された入力音声サンプル(a7)を蓄えるバッファであり、1音声符号化フレーム(20ms)毎に音声サンプル(160サンプル)を取り込み、音声符号化処理部へ(b7)として出力する。以下では1音声符号化フレーム毎に実行される処理について説明する。

【0040】ゲイン計算器(112)は(b7)のレベル情報で

あるRMS値の対数を計算し、その結果である(c7)を出力する。第1の量子化器(以下、「量子化器1」という)(113)は(c7)を5ビットで線形量子化し、その結果である(d7)をビットパッキング器(125)へ出力する。線形予測分析器(114)は、(b7)をDurbin-Levinson法を用いて線形予測分析し、スペクトル包絡情報である10次の線形予測係数(e7)を出力する。LSF係数計算器(115)は、10次の線形予測係数(e7)を10次のLSF(Line Spectrum Frequencies)係数(f7)に変換する。第2の量子化器(以下、「量子化器2」という)(116)は10次のLSF係数(f7)を段数4の多段ベクトル量子化により25ビットで量子化し、その結果であるLSFパラメータインデックス(g7)をビットパッキング器(125)へ出力する。

【0041】ローパスフィルタ(LPF)(120)は(b7)をカットオフ周波数1000Hzでフィルタリングし、(k7)を出力する。ピッチ検出器(121)は、(k7)からピッチ周期を求め、(m7)として出力する。ピッチ周期は正規化自己相関関数が最大となる遅延量として与えられるが、この時の正規化自己相関関数の最大値(17)も出力される。正規化自己相関関数の最大値の大きさは、入力信号(b7)の周期性の強さを表す情報であり、非周期フラグ発生器(122)(後で説明する)で用いられる。また正規化自己相関関数の最大値(17)は、相関関数補正器(119)(後で説明する)で補正された後、有声／無声判定器(126)における有声／無声判定に用いられる。そこでは、補正後の正規化自己相関関数の最大値(j7)が所定の閾値(例えば、0.6)以下であれば無声、そうでなければ有声と判定され、その結果である有声／無声フラグ(s7)が出力される。

【0042】第3の量子化器(以下、「量子化器3」という)(123)はピッチ周期(m7)を入力し対数変換した後、99レベルで線形量子化し、その結果であるピッチインデックス(o7)を周期／非周期ピッチおよび有声／無声情報コード生成器(127)へ出力する。図3に量子化器3(123)への入力であるピッチ周期(20～160サンプルの範囲をとる)とその出力であるインデックスの値(0～98の範囲をとる)の関係を示す。非周期フラグ発生器(122)は、正規化自己相関関数の最大値(17)を入力し、所定の閾値(例えば、0.5)より小さければ非周期フラグを0Nにセット、そうでなければOFFにセットして、非周期フラグ(1ビット)(n7)を非周期ピッチインデックス生成器(124)および、周期／非周期ピッチおよび有声／無声情報コード生成器(127)へ出力する。ここで、非周期フラグ(n7)がONであれば、現フレームが非周期性をもつ音源であることを意味する。

【0043】LPC分析フィルタ(117)は10次の線形予測係数(r7)を係数として用いる全零型フィルタであり、入力音声(b7)からスペクトル包絡情報を除去し、その結果である残差信号(h7)を出力する。ピーキネス計算器(118)は、残差信号(h7)を入力し、ピーキネス値を計算し

(i7)として出力する。このピーキネス値はMELP方式で説明したのと同様の方法を用いて計算する。相関関数補正器(119)は、ピーキネス値(i7)が所定の値(例えば、1.34)より大きければ、正規化自己相関関数の最大値(17)を1.0(有声を示す)にセットし(j7)を出力する。また、前記以下の場合には、前記(17)をそのまま出力する。

【0044】上に述べたピーキネス値の計算および相関関数補正処理は、非周期的なパルスを含むフレームおよび破裂音フレームを検出し、正規化自己相関関数の最大値を1.0(有声を示す値)に補正するための処理である。破裂音フレームでは、部分的にスパイク(鋭いピーク)を持つが、その他の部分は、白色雑音に近い性質の信号になっているため、補正される前の正規化自己相関関数は0.5より小さくなる可能性が大きい(つまり、非周期フラグがONにセットされている可能性が大きい)。一方、ピーキネス値は大きくなる。従って、ピーキネス値により破裂音フレームを検出して正規化自己相関関数を1.0に補正すると、その後の有声/無声判定器(126)における有声/無声判定において有声と判定され、復号の際に非周期パルスが音源として用いられることになるため、破裂音フレームの音質は改善される。また、過渡部に多くみられる非周期的なパルス列を持つフレームも同じ理由により音質は改善される。

【0045】非周期ピッチインデックス生成器(124)は、非周期フレームにおけるピッチ周期(n7)を28レベルで不均一量子化しインデックス(p7)を出力する。この処理内容について説明する。まず、有声/無声フラグ(s7)が有声、かつ、非周期フラグ(n7)がONになっているフレーム(過渡部または破裂音フレームに対応する)に対し、ピッチ周期の度数を調べた結果を図4に、その累積度数を図5に示す。これらは男女各4名(6音声サンプル/各1名)で構成される合計112.12[s](5606フレー

ム)の音声データについて測定した結果である。上記の条件(有声/無声フラグ(s7)が有声、かつ、非周期フラグ(n7)がON)を満たすフレームは、5606フレーム中425フレーム存在した。図4より、その条件を満たすフレーム(以後、非周期フレームと記す)におけるピッチ周期の分布はおよそ25~100に集中していることが分かる。

よって、度数(出現頻度)に基づく不均一量子化を行えば、すなわち、度数が大きなピッチ周期ほど細かく、それが小さいピッチ周期ほど荒く量子化すれば高能率に伝送できる。また、復号器では、非周期フレームのピッチ周期は次式により計算される。

非周期フレームのピッチ周期=伝送されたピッチ周期×
(1.0+0.25×乱数値)

上式で、伝送されたピッチ周期とは、非周期ピッチインデックス生成器(124)の出力であるインデックスにより伝送されるピッチ周期であり、(1.0+0.25×乱数値)を乗算することによりピッチ周期毎にジッタが付加される。したがって、ピッチ周期が大きいほど、ジッタの量も大きくなるため、荒い量子化が許される。

【0046】上記の考えに基づいた非周期フレームのピッチ周期に対する量子化テーブルの例を表4に示す。同表では、入力ピッチ周期が20~24の範囲を1レベル、25~50の範囲を13レベル(2ステップ幅)、51~95の範囲を9レベル(5ステップ幅)、96~135の範囲を4レベル(10ステップ幅)、136~160の範囲を1レベルで量子化し、インデックス(非周期0~27)を出力する。通常のピッチ周期の量子化は、64レベル以上必要であるのに対し、この非周期フレームのピッチ周期の量子化は、度数、復号方法を考慮することにより、28レベルで量子化することが可能となる。

【0047】

【表4】

表4 非周期フレームのピッチ周期に対する量子化テーブル

非周期フレームの ピッチ周期	量子化後の 非周期フレームの ピッチ周期	インデックス	非周期フレームの ピッチ周期	量子化後の 非周期フレームの ピッチ周期	インデックス
20~24	24	非周期 0	51~55	55	非周期 14
25,26	26	非周期 1	56~60	60	非周期 15
27,28	28	非周期 2	61~65	65	非周期 16
29,30	30	非周期 3	66~70	70	非周期 17
31,32	32	非周期 4	71~75	75	非周期 18
33,34	34	非周期 5	76~80	80	非周期 19
35,36	36	非周期 6	81~85	85	非周期 20
37,38	38	非周期 7	86~90	90	非周期 21
39,40	40	非周期 8	91~95	95	非周期 22
41,42	42	非周期 9	96~105	100	非周期 23
43,44	44	非周期 10	106~115	110	非周期 24
45,46	46	非周期 11	116~125	120	非周期 25
47,48	48	非周期 12	126~135	130	非周期 26
49,50	50	非周期 13	136~160	140	非周期 27

【0048】周期/非周期ピッチおよび有声/無声情報コード生成器(127)は、有声/無声フラグ(s7)、非周期フラグ(n7)、ピッチインデックス(o7)、非周期的ピッチ

インデックス(p7)を入力し、7ビット(128レベル)の周期/非周期ピッチ・有声/無声情報コード(t7)を出力する。ここでの処理について以下に述べる。有声/無声

フラグ(s7)が無声を示す場合は、7ビットの符号(128種類の符号語を持つ)のうち、7ビットが全て0の符号語を割り当てる。同フラグが有声を示す場合は、残りの符号語(127種類)を非周期フラグ(n7)に基づき、ピッチインデックス(o7)または非周期ピッチインデックス(p7)に割り当てる。非周期フラグ(n7)がONの時は、非周期ピッチインデックス(p7)(非周期0~27)を7ビット中1ビットおよび2ビットが1となる符号語(28種類)を割り当てる。その他の符号語(99種類)は周期的なピッチインデックス(o7)(周期0~98)に割り当てる。

【0049】以上に基づく周期/非周期ピッチ・有聲/無声情報コードの生成テーブルを表5に示す。通常、伝送誤りにより有聲/無声情報に誤りが発生し、無声フレームが誤って有聲フレームとして復号された場合、周期

表5 周期/非周期ピッチおよび有聲/無声情報コードの生成テーブル

コード	インデックス	コード	インデックス	コード	インデックス	コード	インデックス
0x0	無声	0x20	非周期 15	0x40	非周期 21	0x60	非周期 27
0x1	非周期 0	0x21	非周期 16	0x41	非周期 22	0x61	周期 68
0x2	非周期 1	0x22	非周期 17	0x42	非周期 23	0x62	周期 69
0x3	非周期 2	0x23	周期 16	0x43	周期 42	0x63	周期 70
0x4	非周期 3	0x24	非周期 18	0x44	非周期 24	0x64	周期 71
0x5	非周期 4	0x25	周期 17	0x45	周期 43	0x65	周期 72
0x6	非周期 5	0x26	周期 18	0x46	周期 44	0x66	周期 73
0x7	周期 0	0x27	周期 19	0x47	周期 45	0x67	周期 74
0x8	非周期 6	0x28	非周期 19	0x48	非周期 25	0x68	周期 75
0x9	非周期 7	0x29	周期 20	0x49	周期 46	0x69	周期 76
0xA	非周期 8	0x2A	周期 21	0x4A	周期 47	0x6A	周期 77
0xB	周期 1	0x2B	周期 22	0x4B	周期 48	0x6B	周期 78
0xC	非周期 9	0x2C	周期 23	0x4C	周期 49	0x6C	周期 79
0xD	周期 2	0x2D	周期 24	0x4D	周期 50	0x6D	周期 80
0xE	周期 3	0x2E	周期 25	0x4E	周期 51	0x6E	周期 81
0xF	周期 4	0x2F	周期 26	0x4F	周期 52	0x6F	周期 82
0x10	非周期 10	0x30	非周期 20	0x50	非周期 26	0x70	周期 83
0x11	非周期 11	0x31	周期 27	0x51	周期 53	0x71	周期 84
0x12	非周期 12	0x32	周期 28	0x52	周期 54	0x72	周期 85
0x13	周期 5	0x33	周期 29	0x53	周期 55	0x73	周期 86
0x14	非周期 13	0x34	周期 30	0x54	周期 56	0x74	周期 87
0x15	周期 6	0x35	周期 31	0x55	周期 57	0x75	周期 88
0x16	周期 7	0x36	周期 32	0x56	周期 58	0x76	周期 89
0x17	周期 8	0x37	周期 33	0x57	周期 59	0x77	周期 90
0x18	非周期 14	0x38	周期 34	0x58	周期 60	0x78	周期 91
0x19	周期 9	0x39	周期 35	0x59	周期 61	0x79	周期 92
0x1A	周期 10	0x3A	周期 36	0x5A	周期 62	0x7A	周期 93
0x1B	周期 11	0x3B	周期 37	0x5B	周期 63	0x7B	周期 94
0x1C	周期 12	0x3C	周期 38	0x5C	周期 64	0x7C	周期 95
0x1D	周期 13	0x3D	周期 39	0x5D	周期 65	0x7D	周期 96
0x1E	周期 14	0x3E	周期 40	0x5E	周期 66	0x7E	周期 97
0x1F	周期 15	0x3F	周期 41	0x5F	周期 67	0x7F	周期 98

【0051】ビットパッキング器(125)は、量子化されたRMS値(ゲイン情報)(d7)、LSFパラメータインデックス(g7)、周期/非周期ピッチ・有聲/無声情報コード(t7)を入力すると共に1ビットの同期ビットを付加して、1フレーム(20ms)当たり38ビットの音声情報ビット列(q7)を出力する(表6)。ここに示した実施の形態では音声符号化速度1.9kbpsが実現できる。また、本実施の形態では、MELP方式のようにハーモニック振幅情報は伝送していない。この理由は次の通りである。音声符号化フレーム長を20msと短くしているため(ME

的音源が使用されるため再生音声の品質は著しく劣化する。本発明においては、非周期ピッチインデックス(p7)(非周期0~27)を7ビット中1ビットおよび2ビットが1となる符号語(28種類)に割り当てることにより、無声の符号語(0x0)が伝送誤りにより1または2ビット誤ったとしても、非周期的なピッチパルスにより音源信号が作られるため、伝送誤りによる影響を軽減することが出来る。また、前述したMELP方式では非周期フラグの伝送に1ビット使用していたが、本発明を用いることにより、それが不要となり、伝送ビット数の削減が可能となる。

【0050】

【表5】

LP方式では22.5ms)、LSFパラメータを抽出する周期が短くなり、スペクトル表現の正確さが向上する。従ってハーモニック振幅情報は必要としない。

【0052】

【表6】

表6 本方式のビット配分(1.9kbps)

パラメータ	ビット数
LSFパラメータ	25
ゲイン1回/フレーム	5
周期/非周期ピッチ・有声/無声 情報コード	7
同期ビット	1
合計ビット/20ms フレーム	38

【0053】次に、図2を用いて本発明の音声復号方法が適用された音声復号器の一実施の形態について説明する。図2において、ビット分離器(131)は1フレーム毎に受信した38ビットの音声情報ビット列(a8)を各パラメータ毎に分離し、周期/非周期ピッチ・有声/無声情報コード(b8)、ゲイン情報(i8)およびLSFパラメータインデックス(f8)を出力する。有声/無声情報・ピッチ周期復号器(132)は周期/非周期ピッチ・有声/無声情報コード(b8)を入力し、前記表5に示したテーブルに基づき、無声/周期的/非周期的のうちどれであるかを求め、無声ならば、ピッチ周期(c8)を所定の値(例えば、50)にセット、有声強度(d8)を0にセットして出力する。周期的および非周期的の場合は、ピッチ周期(c8)を

復号処理(非周期的の場合は表4を用いる)して出力し、有声強度(d8)を1.0にセットして出力する。
【0054】ジッタ設定器(133)は、周期/非周期ピッチ・有声/無声情報コード(b8)を入力し、表5のテーブルに基づき、無声/周期的/非周期的のうちどれであるかを求め、無声または非周期的を示す場合は、ジッタ値(e8)を所定の値(例えば、0.25)にセットして出力する。周期的を示す場合は、ジッタ値(e8)を0にセットして出力する。LSF復号器(134)はLSFパラメータインデックス(f8)から10次のLSF係数(g8)を復号し出力する。傾斜補正係数計算器(135)は、10次のLSF係数(g8)から傾斜補正係数(h8)を計算する。ゲイン復号器(136)はゲイン情報(i8)を復号し、ゲイン(j8)を出力する。

【0055】パラメータ補間器(137)は、各パラメータ(c8)、(d8)、(e8)、(g8)、(h8)および(j8)についてそれぞれピッチ周期に同期して線形補間し、(k8)、(n8)、(l8)、(u8)、(v8)および(w8)を出力する。ここでの線形補間処理は、次式により実施される。

補間後のパラメータ=現フレームのパラメータ×int+前フレームのパラメータ×(1.0-int)

ここで、現フレームのパラメータは(c8)、(d8)、(e8)、(g8)、(h8)および(j8)のそれぞれに対応し、補間後のパラメータは(k8)、(n8)、(l8)、(u8)、(v8)および(w8)のそれぞれに対応する。前フレームのパラメータは、前フレームにおける(c8)、(d8)、(e8)、(g8)、(h8)および(j8)を保持しておくことにより与えられる。また、intは補間係数であり、次式で求める。

$int = to / 160$

ここで、160.0は音声復号フレーム長(20ms)当たりの

サンプル数、toは、復号フレームにおける1ピッチ周期の開始点であり、1ピッチ周期分の再生音声は復号される毎にそのピッチ周期が加算されることにより更新される。toが160を超えるとそのフレームの復号処理が終了したことになり、toから160が減算される。ここで、補間係数intを1.0に固定するとピッチ周期に同期した線形補間処理は実施されないことになる。

【0056】ピッチ周期計算器(138)は、補間されたピッチ周期(k8)およびジッタ値(18)を入力し、ピッチ周期(m8)を次式により計算する。

ピッチ周期(m8)=ピッチ周期(k8)×(1.0-ジッタ値(18)×乱数値)

ここで、乱数値は-1.0~1.0の範囲の値をとる。このピッチ周期(m8)は小数を持つが、四捨五入され整数に変換される。整数に変換されたピッチ周期(m8)を以下では”T”と表す。上式より、無声または非周期的フレームではジッタ値が所定の値(この実施の形態では、0.25)にセットされているのでジッタが付加され、完全な周期的フレームではジッタ値が0にセットされているのでジッタは付加されない。但し、ジッタ値はピッチ毎に補間処理されているので、0~0.25の範囲をとるため中間的なジッタ量が付加されるピッチ区間も存在する。このように非周期ピッチ(ジッタが付加されたピッチ)を発生することは、MELP方式の説明で述べたように過渡部、破裂音で生じる不規則な(非周期的な)声門パルスを表現することにより、トーンの雑音を低減する効果がある。

【0057】1ピッチ波形復号器(152)は、ピッチ周期(Tサンプル)毎の再生音声(e9)を復号し出力する。従って、このブロックに含まれる全てのブロックはピッチ周期(T)を入力し、それに同期して動作する。第1の混合音源発生器(以下、「混合音源1発生器」という)(141)は、補間された有声強度(n8)(0~1.0の値をとる)に基づき、次式により単一パルス発生器(139)から出力される単一パルス信号(o8)(Tサンプルの期間内に1つのパルスが発生され、それ以外のサンプル値は0)と雑音発生器(140)から出力される白色雑音(p8)を次式に基づいて合成して、第1の混合音源信号(以下、「混合音源1」という)1(q8)を生成する。ここで、単一パルス信号(o8)および白色雑音(p8)のレベルは所定のRMS値になるように調整されている。

混合音源1(q8)=単一パルス信号(o8)×有声強度(n8)+白色雑音(p8)×(1.0-有声強度(n8))

この処理により、無声音源(白色雑音)から有声音源(単一パルス信号)(またはこの逆)への急激な切り替わりを防ぐことができ、再生音声の品質が改善される。混合音源1(q8)は、完全な有声フレームでは有声強度(n8)が1.0となるため単一パルス信号(o8)に等しくなり、完全な無声フレームでは有声強度(n8)が0となるため白色雑音(p8)に等しくなる。

【0058】線形予測係数計算器(147)は補間された10次のL S F係数(u8)から線形予測係数(x8)を計算する。スペクトル包絡形状計算器(146)は、線形予測係数(x8)から再生音声のスペクトル包絡形状情報(y8)を求め出力する。この処理の実現例について説明する。線形予測係数(x8)をTポイントD F T(離散フーリエ変換)を行い、その大きさ(マグニチュード)を計算することによりL P C分析フィルタの伝達関数を求めた後、それを反転させることにより逆特性(これはL P C合成フィルタの伝達関数、すなわち再生音声のスペクトル包絡形状に対応)を求める。それを正規化しスペクトル包絡形状情報(y8)として出力する。スペクトル包絡形状情報(y8)は図6(a)に示すような0~4000Hzの再生音声のスペクトル包絡成分を、Tが偶数の時は $T/2$ 個のD F T係数、Tが奇数の時は $(T-1)/2$ 個のD F T係数で表す情報である。

【0059】混合音源用フィルタリング器(142)は、混合音源1(q8)を入力し、TポイントD F Tを行い、Tが偶数の時は $T/2$ 個のD F T係数、Tが奇数の時は $(T-1)/2$ 個のD F T係数を求める(図6(b))。但し、図6(b)では、簡単のため混合音源1(q8)が単一パルスである場合(完全に有聲の場合)について示しており、各D F T係数は全て1.0の値を持つ。次に、スペクトル包絡形状情報(y8)と閾値(f9)を入力し、図6

(a)に示すようにスペクトル包絡形状情報(y8)を表すD F T係数の大きさが閾値以上になる周波数領域(同図ではa~bおよびc~dの区間)(以後、有聲領域と呼ぶ)を調べ、混合音源1(q8)のD F T結果(図6(b))において、有聲領域以外の領域(以後、無聲領域と呼ぶ)のD F T係数を0にしたD F T係数列(r8)(図6(c)の実線)を出力する。なお、この閾値の値としては、0.6~0.9の範囲の値が適当であり、ここでは閾値は0.8とする。また、混合音源1(q8)のD F T結果(図6

(b))において、有聲領域のD F T係数を0としたD F T係数列(s8)(図6(c)の点線)を出力する。つまり、ここでは混合音源1(q8)を、スペクトル包絡形状情報(y8)の大きさが閾値以上になる周波数領域(有聲領域)の成分と、それ以外の周波数領域(無聲領域)の成分に分けて、D F T係数の形で、それぞれ(r8)、(s8)として出力している。

【0060】雑音音源用フィルタリング器(143)は、白色雑音(p8)を入力し、TポイントD F Tを行い、Tが偶数の時は $T/2$ 個のD F T係数、Tが奇数の時は $(T-1)/2$ 個のD F T係数を求める(図6(d))。次に、スペクトル包絡形状情報(y8)と閾値(f9)を入力し、図6(a)に示すようにスペクトル包絡形状情報(y8)を表すD F T係数の大きさが閾値以上になる周波数領域(有聲領域)を調べ、白色雑音(p8)のD F T結果(図6(d))において、有聲領域のD F T係数を0にしたD F T係数列(t8)(図6(e))を出力する。第2の混合

音源発生器(以下、混合音源2発生器)(144)は、D F T係数列(s8)(図6(c)の点線)およびD F T係数列(t8)(図6(e))を入力し、所定の割合でこれらを混合して、その結果であるD F T係数列(z8)を出力する。本実施形態では、D F T係数列(s8)を6割、D F T係数列(t8)を4割で混合している。なお、この混合比率は、(s8)を5~7割、(t8)が5~3割の範囲の値とするのが好適である。第3の混合音源発生器(以下、「混合音源3発生器」という)(145)は、D F T係数列(r8)およびD F T係数列(z8)を入力し、これらを加算した後(この時のD F T係数を図6(f)に示す)、I D F T(逆離散フーリエ変換)を行い時間波形に戻し、混合音源信号(g9)として出力する。

【0061】完全な無声フレームでは、有聲強度(n8)が0なので、混合音源1(q8)および混合音源信号(g9)は白色雑音(p8)と等しくなる。従って、上記の混合音源信号(g9)を生成する処理を実施する前に、切替え器(153)は有聲強度(n8)を観測し、それが0(完全無音)であれば混合音源信号として白色雑音(p8)を選択し、その他は混合音源信号(g9)を選択し(a9)として出力する。これにより完全無音フレームでの処理量を節約することができる。

【0062】以下に、スペクトル包絡形状計算器(146)、混合音源用フィルタリング器(142)、雑音音源用フィルタリング器(143)、混合音源2発生器(144)および混合音源3発生器(145)を用いた混合音源信号生成の効果について説明する。入力音声に対しスペクトル包絡形状を求め、その大きさ(マグニチュード)が閾値以上になる周波数成分と、閾値より小さくなる周波数成分に分割し、それぞれの時間波形についてピッチ周期を遅延時間としたときの正規化自己相関関数を計算し、度数を調べた結果を図7に、その累積度数を図8に示す。但し、この測定では有聲フレーム(周期的および非周期的フレーム)のみを有効とした。被験音声として男女各4名(2音声サンプル/各1名)で構成される合計36.22[s](1811フレーム)の音声データを用いたが、有効フレーム(有聲フレーム)数は、1616フレームであった。また、これらは閾値を0.8とした時の測定例である。

【0063】図7、8より、スペクトル包絡形状の大きさ(マグニチュード)が閾値以上になる成分では、正規化自己相関関数の分布は1.0(最大値)付近に集中し、閾値より小さくなる成分では0.25付近で最大となり、広く分布することが分かる。正規化自己相関関数が大きい程、入力音声の周期性は強く、それが小さい程、入力音声の周期性は弱い(白色雑音に近くなる)ことを意味する。従って、本発明の復号方法のようにスペクトル包絡形状の大きさ(マグニチュード)が閾値より小さい周波数領域のみに対し、白色雑音を付加して混合音源とする方法が妥当であることが分かる。この処理により、MELP方式のように帯域毎の有声性情報を伝送する必要な

しに前述のLPC方式の問題点Aであるbuzz音を低減することができる。

【0064】また、参考文献[6]（線形予測分析・合成方式の復号器）において提案された方法では、LPC方式の問題点A（buzz音）は軽減できるものの、再生音声の音質が雑音的になるという欠点があると前述したが、その理由は次の通りである。図8において、スペクトル包絡形状の大きさ（マグニチュード）が閾値より小さい周波数領域の成分（○により表示）では、正規化自己相関関数が0.6以上に分布する割合が全体の約20%を占めている。従って、全てのフレームにおいてスペクトル包絡形状の大きさ（マグニチュード）が閾値より小さい周波数領域を白色雑音で置き換えてしまうと、再生音声の雑音感が増大し、品質を劣下させてしまう。本発明の方法を用いればこの問題を解決することができる。

【0065】適応スペクトルエンハンスメントフィルタ(148)は、線形予測係数(x8)に帯域幅拡張処理を施したものを係数とする適応極／零フィルタであり、表2の示した通り、ホルマントの共振を鋭くし、自然音声のホルマントに対する近似度を改善することにより再生音声の自然性を向上させる。さらに、補間された傾斜補正係数(v8)を用いてスペクトルの傾きを補正して音のこもりを低減する。前記切替え器(153)の出力(a9)は適応スペクトルエンハンスメントフィルタ(148)によりフィルタリングされ音源信号(b9)が出力される。LPC合成フィルタ(149)は、線形予測係数(x8)を係数として用いる全極型フィルタであり、音源信号(b9)に対しスペクトル包絡情報を付加して、その結果である信号(c9)を出力する。ゲイン調整器(150)は(c9)に対しゲイン情報(w8)を用いてゲイン調整を行い、(d9)を出力する。パルス拡散フィルタ(151)は、自然音声の声門パルス波形に対するパルス音源波形の近似度を改善するためのフィルタであり、(d9)をフィルタリングして自然性が改善された再生音声(e9)を出力する。このパルス拡散フィルタの効果は表2の示す通りである。

【0066】なお、以上説明した本発明の音声符号化装置および音声復号装置は、DSP（デジタル・シグナル・プロセッサ）によって容易に実現可能である。また、前述した本発明の音声復号方法は、音声符号化器として従来方式（LPC方式）のものを用いた場合にも、そのまま適用することができる。さらに、上述の量子化レベル数、符号語のビット数、音声符号化フレーム長、線形予測係数、LSF係数などの次数、各フィルタのカットオフ周波数などは、それぞれ前述した実施の形態において用いられた値に限られることはなく、それぞれの場合に応じた値を採用することができる。

【0067】

【発明の効果】以上説明したように、本発明の音声符号化復号方法および装置を用いることにより、付加的な情報ビットを伝送すること無く、従来方式（LPC）にお

いて品質劣下の原因となっているbuzz音、トーン的雑音を低減でき、再生音声の音質を向上できると共に、従来方式（MELP）よりも符号化速度を下げる事が可能となる。したがって、無線通信に用いる際には、周波数利用効率の向上を図ることができる。

【図面の簡単な説明】

【図1】 本発明の音声符号化方法が適用された音声符号化器の一実施の形態の構成を示すブロック図である。

【図2】 本発明の音声復号方法が適用された音声復号器の一実施の形態の構成を示すブロック図である。

【図3】 ピッチ周期とインデックスの関係を説明するための図である。

【図4】 ピッチ周期の度数を説明するための図である。

【図5】 ピッチ周期の累積度数を説明するための図である。

【図6】 本発明の復号方法における混合音源作成方法を説明するための図である。

【図7】 正規化自己相関関数の度数を説明するための図である。

【図8】 正規化自己相関関数の累積度数を説明するための図である。

【図9】 従来方式（LPC）の音声符号化器の構成を示す図である。

【図10】 従来方式（LPC）の音声復号器の構成を示す図である。

【図11】 LPC方式およびMELP方式のスペクトルについて説明するための図である。

【図12】 従来方式（MELP）の音声符号化器の構成を示す図である。

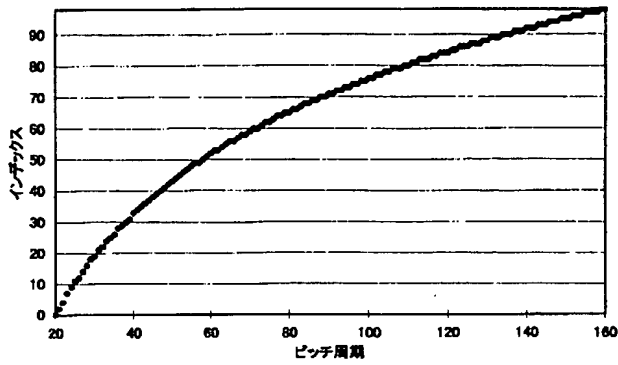
【図13】 従来方式（MELP）の音声復号器の構成を示す図である。

【符号の説明】

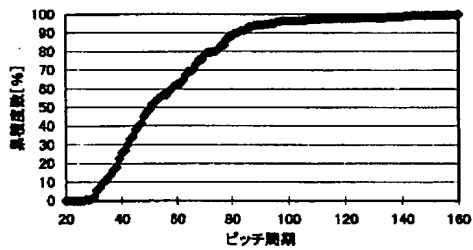
111 フレーム化器、112 ゲイン計算機、113 量子化器1、114 線形予測分析器、115 LSF係数計算器、116 量子化器2、117 LPC分析フィルタ、118 ピーケネス計算器、119 相関関数補正器、120 ローパスフィルタ、121 ピッチ検出器、122 非周期フラグ発生器、123 量子化器3、124 非周期ピッチインデックス生成器、125 ビットパッキング器、126 有声／無声判定器、127 周期／非周期ピッチおよび有声／無声情報コード生成器、131 ビット分離器、132 有声／無声情報・ピッチ周期復号器、133 ジッタ設定器、134 LSF復号器、135 傾斜補正係数計算器、136 ゲイン復号器、137 パラメータ補間器、138 ピッチ周期計算器、139 単一パルス発生器、140 雑音発生器、141 混合音源1発生器、142 混合音源用フィルタリング器、143 雑音音源用フィルタリング器、144 混合音源2発生器、145 混合

音源3発生器、146 スペクトル包絡形状計算器、147 線形予測係数計算器、148 適応スペクトルフィルタ、149 LPC合成フィルタ、150 ゲイン

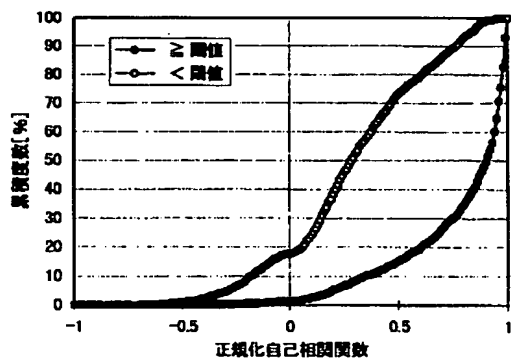
【図3】



【図5】

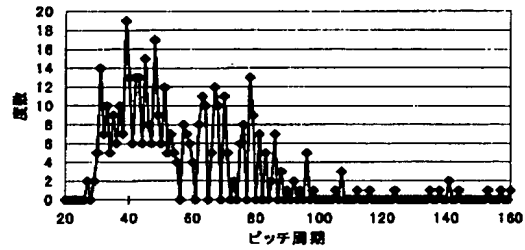


【図8】

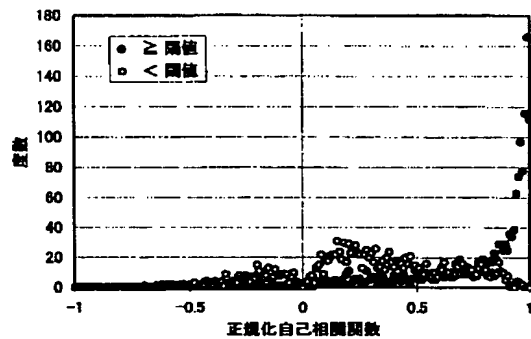


調整器、151 パルス拡散フィルタ、152 1ピッチ波形復号器

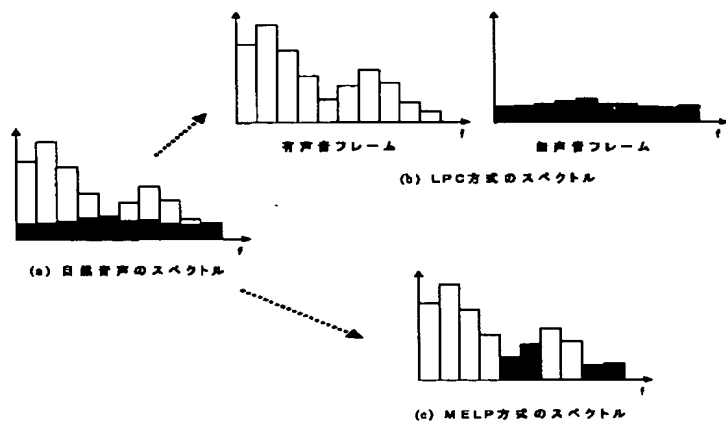
【図4】



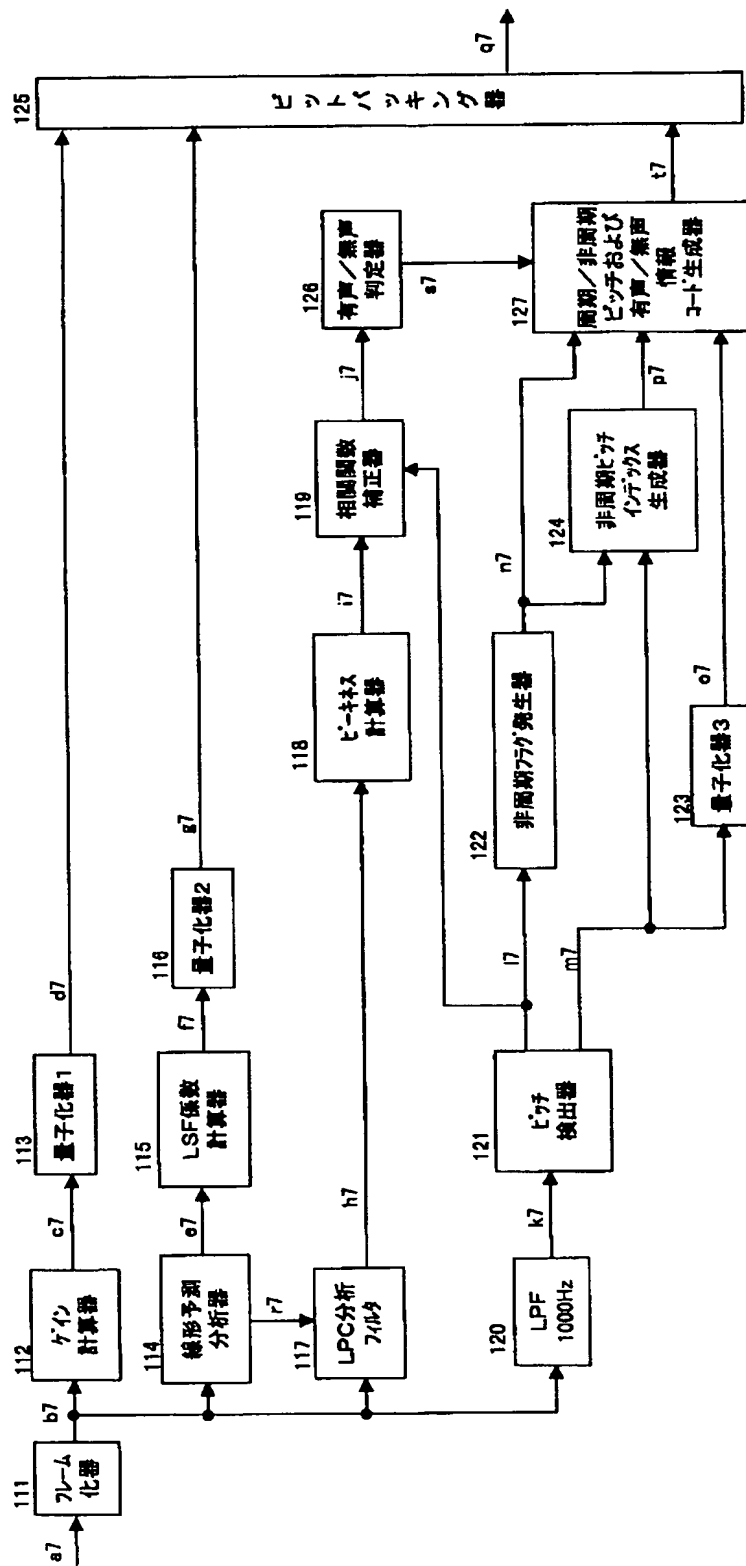
【図7】



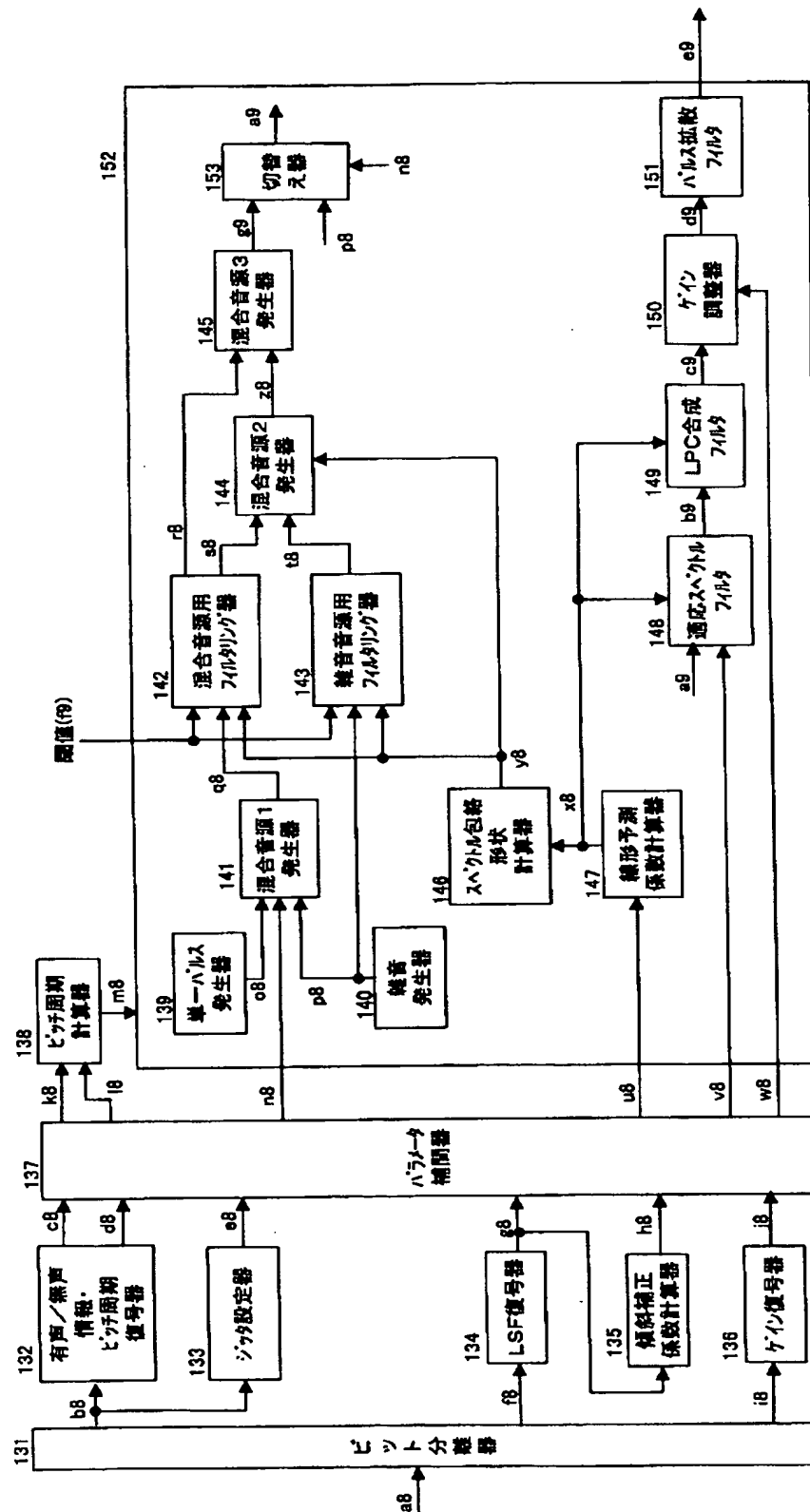
【図11】



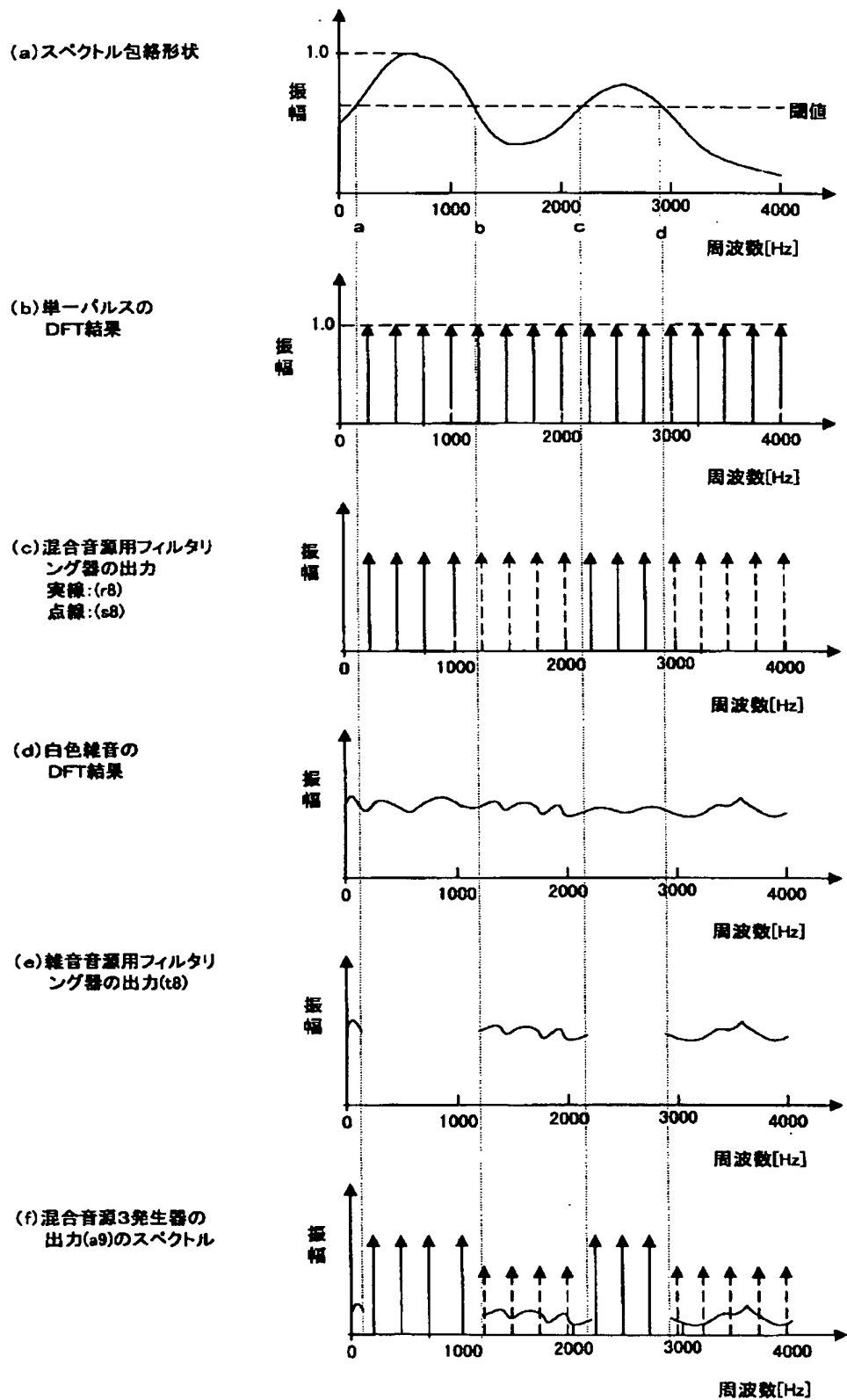
【図1】



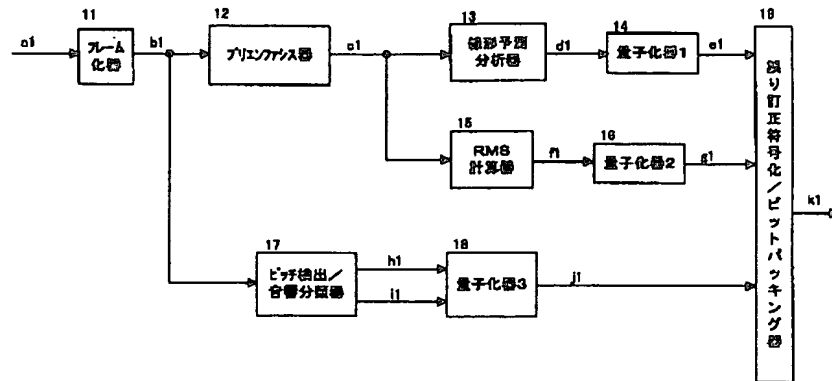
【图 2】



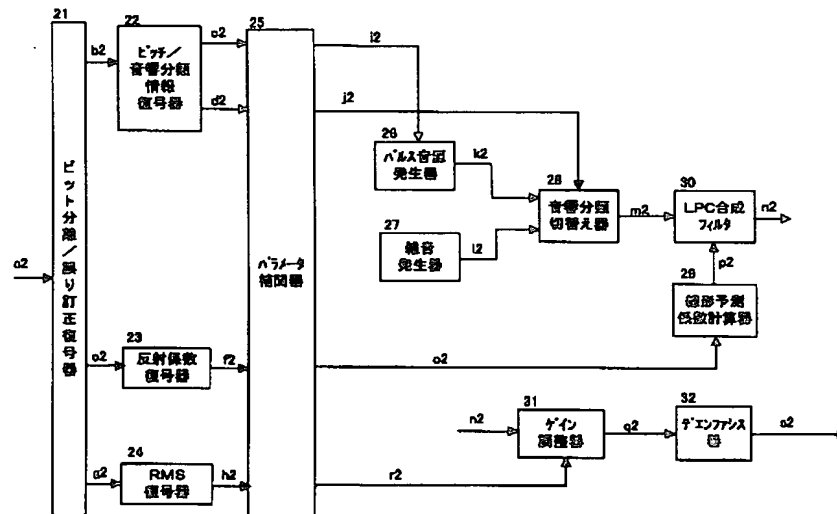
【図 6】



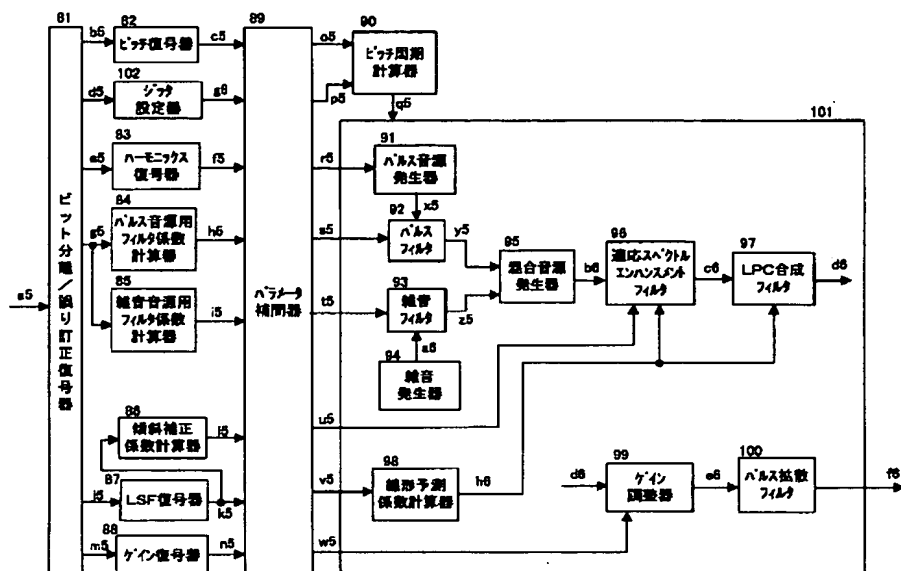
【図 9】



【図 10】



【図 13】



フロントページの続き

Fターム(参考) 5D045 CA04

5J064 AA01 BB03 BB04 BB12 BC08

BC12 BC16 BC22 BC27 BD02

9A001 BB03 BB04 CC05 EE04 GG05

GG22 HH15 HH16 HH17 HH18

KK37 LL02